

Cosmogenic Radionuclides

Physics of Earth and Space Environments

The series *Physics of Earth and Space Environments* is devoted to monograph texts dealing with all aspects of atmospheric, hydrospheric and space science research and advanced teaching. The presentations will be both qualitative as well as quantitative, with strong emphasis on the underlying (geo)physical sciences. Of particular interest are

- contributions which relate fundamental research in the aforementioned fields to present and developing environmental issues viewed broadly
- concise accounts of newly emerging important topics that are embedded in a broader framework in order to provide quick but readable access of new material to a larger audience

The books forming this collection will be of importance for graduate students and active researchers alike.

Series Editors:

Rodolfo Guzzi
Responsabile di Scienze della Terra
Head of Earth Sciences
Via di Villa Grazioli, 23
00198 Roma, Italy

Louis J. Lanzerotti
Jersey Institute of Technology
Newark, NJ
USA

Ulrich Platt
Ruprecht-Karls-Universität Heidelberg
Institut für Umweltphysik
Im Neuenheimer Feld 229
69120 Heidelberg, Germany

Jürg Beer • Ken McCracken • Rudolf von Steiger

Cosmogenic Radionuclides

Theory and Applications in the Terrestrial
and Space Environments

With 196 Figures and 9 Tables

 Springer

Jürg Beer
Eawag
Überlandstr. 133
8600 Dübendorf
Switzerland
beer@eawag.ch

Ken McCracken
IPST, University of Maryland, USA
Jellore Lane
2575 High Range NSW
Australia
jellore@hinet.net.au

Rudolf von Steiger
International Space Science Institute
Hallerstrasse 6
3012 Bern
Switzerland
vsteiger@issibern.ch

ISSN 1610-1677 e-ISSN 1865-0678
ISBN 978-3-642-14650-3 e-ISBN 978-3-642-14651-0
DOI 10.1007/978-3-642-14651-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011939486

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Acknowledgements

It has taken eight years to write this book. When we started there was little interaction between the cosmogenic and cosmic ray research communities. As the book was written, that situation was changing rapidly. Many of our colleagues provided key insights, ideas and publications that hastened that change. In particular, the motivation towards close interaction came from both the cosmogenic and cosmic ray communities. Each community could see that the whole would be much greater than the sum of the two parts. We acknowledge this widespread contribution to the present day study and utilization of the cosmogenic radionuclides.

One of the most significant contributions was the pioneering work of Josef Masarik in providing the quantitative links between the cosmic radiation and the production of the cosmogenic radionuclides in the atmosphere. Many people contributed to the provision of the century and millennia-long ice cores that are fundamental to all studies and applications in this book; we acknowledge in particular the initiative and the work of Hans Oeschger, Chester Langway and Willi Dansgaard. The radiocarbon tree ring record for the past 10,000 years plays also a vital role, and we acknowledge the decade long work of Hans Suess and Minze Stuiver that has yielded the present day archive. Willy Wöflfi, Martin Suter and Georges Bonani designed and built the AMS system resulting in the analysis of more than 15,000 samples. These samples have been processed over the years by Christian Obrist, Caroline Stengel, Silvia Bollhalder, and Irene Brunner.

In particular we acknowledge the contributions of the late Bernhard Lehmann who was to be a co-author of this text, and to whose memory the book is dedicated.

The authors have benefited greatly from the work and stimulus of their colleagues in Switzerland, the USA, Australia, South Africa, and elsewhere. We acknowledge the work of Ulrich Siegenthaler, Robert Finkel, Stephan Baumgartner, Irka Hajdas, Gerhard Wagner, Raimund Muscheler, Maura Vonmoos, Ulla Heikkilä, Friedhelm Steinhilber, and Jose Abreu in drawing together the several data sources, and developing the means to analyze them with precision. The cosmic ray community, notably Frank McDonald, Harm Moraal, Bill Webber, Rogelio Caballero-Lopez,

Don Reames, Peggy Shea, Don Smart, Gisela Dresschoff, Marc Duldig and John Humble have all assisted us with ideas, criticism and above all, encouragement.

Several colleagues have significantly improved the manuscript by checking individual chapters: Marcus Christl, Alexander Blinov, Johannes Feichter, Ulla Heikkilä, John Humble, Jozef Masarik, and Don Reames.

We gratefully acknowledge Yvonne Lehnhard, Lydia Zweifel and Irene Brunner who worked hard to turn our figures from the literature and our rough sketches into color figures.

A key role in the development of this book was played by the International Space Science Institute of Bern, Switzerland. Their enthusiastic support over the years made possible the international collaboration without which this book could not have been written.

Contents

Part I Introduction

1 Motivation	3
2 Goals	7
Reference	9
3 Setting the Stage and Outline	11

Part II Cosmic Radiation

4 Introduction to Cosmic Radiation	17
5 The Cosmic Radiation Near Earth	19
5.1 Introduction and History of Cosmic Ray Research	19
5.2 The “Rosetta Stone” of Paleocosmic Ray Studies	21
5.3 Some Important Definitions	22
5.4 The Origin and Properties of the Galactic Cosmic Radiation	27
5.5 Our Variable Sun	33
5.6 The Heliosphere, the Termination Shock, and the Current Sheet	41
5.7 Modulation of the Cosmic Radiation in the Heliosphere	44
5.7.1 The Cosmic Ray Propagation Equation	45
5.7.2 The Local Interstellar Spectrum	48
5.7.3 The Cosmic Ray Modulation Function and Potential	51
5.7.4 Practical Applications of the Modulation Function	59
5.7.5 Drift Effects (qA Positive and qA Negative Effects)	60
5.7.6 Shock Wave Effects (The Forbush Decrease and GMIRs)	62

5.8	Geomagnetic Field Effects	64
5.8.1	The Properties of the Geomagnetic Field	64
5.8.2	The Geomagnetic Cut-off Rigidity	68
5.8.3	The Earth's Magnetosphere and the Polar Aurora	73
	References	77
6	Instrumental Measurements of the Cosmic Radiation	79
6.1	Introduction	79
6.2	Ionization Chambers and Muon Telescopes	80
6.3	The IGY and IQSY Neutron Monitors, and Spaceship Earth	83
6.4	Satellite Borne Detectors	88
6.5	Latitude Effects and the Yield Functions	90
6.6	Inter-calibration of the Different Cosmic Ray Records	93
6.7	Cosmic Ray Archives	96
	References	97
7	Time Variations of the Cosmic Radiation	99
7.1	Introduction and Atmospheric Effects	99
7.2	The Eleven- and Twenty-Two-Year Variations	100
7.3	The Long-term Variations	103
7.4	Forbush Decreases, Globally Merged Interaction Regions and Some Smaller Effects	106
	References	109
8	The Solar Cosmic Radiation	111
8.1	Historical Overview	111
8.2	The Observed Production of Cosmic Rays by the Sun	112
8.2.1	Ground Level Events	112
8.2.2	SEP Events Observed by Satellites	115
8.2.3	Paleo-Cosmic Ray Measurements of SEP Events	119
8.3	Overall Characteristics of the Solar Cosmic Radiation	125
8.3.1	The Energy Spectra	125
8.3.2	The Effect of Longitude Relative to the Central Solar Meridian	127
8.3.3	The Frequency of Occurrence, and the Detection of Historic SEP Events	128
	References	130

Part III Cosmogenic Radionuclides

9	Introduction to Cosmogenic Radionuclides	135
10	Production of Cosmogenic Radionuclides in the Atmosphere	139
10.1	Introduction	139

10.2 Interaction of Primary Cosmic Rays with the Atmosphere	142
10.2.1 Production of Secondary Particles	142
10.2.2 Ionization and Excitation Processes	148
10.2.3 Simulated Atmospheric Proton and Neutron Fluxes	150
10.3 Production of Cosmogenic Radionuclides in the Atmosphere ...	157
10.3.1 Early Production Models	159
10.3.2 Production Cross-Sections	161
10.3.3 Production Rates and Inventories	163
10.4 Production Results and Analytical Tools	172
References	176
11 Production of Cosmogenic Radionuclides in Other	
Environmental Systems	179
11.1 Introduction	179
11.2 Terrestrial Solid Matter (Rocks, Ice)	182
11.2.1 ^{36}Cl Production in Limestone and Dolomite	183
11.2.2 ^{10}Be and ^{14}C Production in Ice	185
11.3 Extraterrestrial Solid Matter	186
References	189
12 Alternative Production Mechanisms	191
12.1 Introduction	191
12.2 Natural Production Mechanisms	191
12.2.1 Cosmic Ray Induced Reactions	191
12.2.2 Radioactive Decay-Induced Reactions	195
12.3 Anthropogenic Production Mechanisms	198
12.3.1 Nuclear Power Plant and Nuclear Bomb-Induced Reactions	198
12.3.2 Research, Industrial, and Medical Induced Reactions	200
References	201
13 Transport and Deposition	203
13.1 Introduction	203
13.2 Basics of the Atmosphere	205
13.3 Removal or Scavenging Processes	211
13.3.1 Wet Deposition	211
13.3.2 Dry Deposition	214
13.3.3 Gravitational Settling	214
13.3.4 The Big Picture	215
13.4 Modelling the Atmospheric Transport	216
13.4.1 Summary	222
13.5 Geochemical Cycles	223
13.5.1 Introduction	223
13.5.2 The Beryllium Cycle	223
13.5.3 Carbon Cycle	225

13.5.4 The Chlorine Cycle	236
13.5.5 The Iodine Cycle	238
References	238
14 Archives	241
14.1 Introduction	241
14.2 Intrinsic Properties of the Cosmogenic Radionuclide Archives	242
14.3 Time Scales	244
14.4 Examples of Archives	248
14.5 Proxies and Surrogates	258
14.6 Properties of Data in the Cosmogenic Archives	260
14.6.1 Sampling Effects	260
14.6.2 Transfer Functions	262
14.7 Modelled Transfer Functions	267
14.7.1 ^{10}Be and ^7Be in the Atmosphere	267
14.7.2 ^{10}Be and ^{26}Al in Deep-Sea Sediments	270
References	276
15 Detection	279
15.1 Introduction	279
15.2 Low-Level Decay Counting	280
15.3 Accelerator Mass Spectrometry	282
15.4 Decay Versus Atom Counting	287
15.5 Other Techniques, Optical Methods	289
15.5.1 Final Remarks	290
References	290
 Part IV Applications	
16 Introduction to Applications	295
17 Solar Physics	297
17.1 Introduction	297
17.2 Solar Periodicities and the “Grand Minima” in the Cosmogenic Radionuclide Record	298
17.2.1 Solar Periodicities: Time Domain Studies	298
17.2.2 Solar Periodicities: Frequency Domain Studies	303
17.3 Cosmic Ray and Solar Effects in the Past	310
17.3.1 The Past Millennium	310
17.3.2 The Past 10,000 Years (the “Holocene”)	312
17.3.3 The Long Solar Minimum of 2007–2009	314
17.4 The Heliomagnetic Field Throughout the Past 10,000 Years	316
17.5 Solar Irradiance and Terrestrial Climate	320
17.6 Radiation Doses on Earth and in Space in the Future	325

17.7	Quantitative Measures of Solar Activity for the Past	325
17.7.1	Reconstructed Sunspot Numbers	326
17.7.2	Modulation Function	327
	References	327
18	Galactic Astronomy	331
18.1	Introduction	331
18.2	Galactic Structure	332
18.3	Individual Supernova	336
	References	339
19	Atmosphere	341
19.1	Introduction	341
19.2	Studies of Atmospheric Mixing	342
19.3	^{36}Cl Bomb Pulse as a Tracer of Atmospheric Transport	347
19.4	Concentrations and Fluxes	350
	References	353
20	Hydrosphere	355
20.1	Introduction	355
20.2	Tritium	358
20.3	Carbon-14	359
20.4	Krypton-81	360
20.5	Chlorine-36	362
20.6	Beryllium-7 to Beryllium-10 Ratio	365
	References	367
21	Geosphere	369
21.1	Introduction	369
21.2	Geomagnetic Field Intensity	371
21.3	Transport of Cosmogenic Radionuclides in Geological Systems ...	377
21.3.1	Introduction	377
21.3.2	Migration in Ice	378
21.3.3	Transport in Soils	380
21.3.4	Transport in Rocks	384
21.3.5	Formation of Loess Plateaus	384
21.3.6	Subduction	386
	References	387
22	Biosphere	389
22.1	Introduction	389
22.2	Radiocarbon Applications	390
22.3	Chlorine-36 in Ecosystems	393
22.4	Iodine-129	394
22.5	Aluminium-26	394
	References	395

23	Dating	397
23.1	Introduction	397
23.2	Absolute Dating	399
23.2.1	Principle of Radiocarbon Dating	401
23.2.2	Exposure Dating	406
23.2.3	$^{10}\text{Be}/^{36}\text{Cl}$ - and $^7\text{Be}/^{10}\text{Be}$ -Dating	411
23.3	Synchronization of Records	414
23.3.1	^{10}Be or ^{36}Cl with ^{14}C During the Holocene	415
23.3.2	The Use of Time Markers	416
	References	417
	Glossary	419
	Index	423

List of Boxes

Box 5.3.1	Relativistic Physics	22
Box 5.5.1	The Sun and the Sunspot Number	34
Box 5.7.2.1	Estimates of the Local Interstellar Spectrum (LIS)	50
Box 5.7.3.1	Modulation of the Alpha and Heavier Components of the Cosmic Radiation	55
Box 10.1.1	The Earth's Atmosphere	140
Box 10.2.1.1	Cross-sections	143
Box 10.2.1.2	Mean Free Path	147
Box 10.2.3.1	Geomagnetic Modulation of Atmospheric Neutron and Proton Fluxes	156
Box 10.3.1.1	Nuclear Reactions	160
Box 10.3.3.1	Treating Non-proton Particles	164
Box 11.1.1	Production of a Cosmogenic Radionuclide in Solid Matter	180
Box 12.2.1.1	Muons and Neutrinos	192
Box 13.2.1	Potential Temperature	208
Box 13.5.3.3.1	Mathematical Treatment of Sensitivity of the Carbon Cycle to Internal Factors	233
Box 14.6.2.1	Solving Transfer Functions in Box Models	264
Box 15.3.1	Sample Preparation for AMS	286
Box 17.2.1.1	Period, Frequency, Harmonics, and Frequency Domain Analysis	298
Box 17.2.2.1	Signals, Noise, Numerical Filters, and Statistics	304
Box 18.2.1	The Anomalous Cosmic Radiation	335
Box 21.3.3.1	Inventories, Age, and Erosion Rates	382
Box 23.2.1.1	Isotopic Fractionation	401

Abbreviations

ACR	Anomalous cosmic radiation
asl	Above sea level
AU	Astronomical unit
AMS	Accelerator mass spectrometry
B/M	Brunhes/Matuyama
CERN	The European Organization for Nuclear Research
CME	Coronal mass ejection
CMP	Central meridian passage
DNA	Deoxyribonucleic acid
DYE 3	Ice core drill site in South-Greenland
DOME C	Ice core drill site within EPICA
ECHAM	European Centre/Hamburg model
EPICA	European project for ice coring in Antarctica
FLUKA	Nuclear interaction software
GCM	General circulation model
GCR	Galactic cosmic rays
GEANT	Nuclear interaction software
GISP	Greenland ice sheet project
GLE	Ground level event
GMIR	Global merged interaction regions
GRIP	Greenland ice core project
HMF	Heliomagnetic field
HTO	“Tritiated water”, $^3\text{H}_2\text{O}$
IGY	International geophysical year
IMP	Interplanetary monitoring platform
INTCAL04	Radiocarbon age calibration data 2004
IPCC	Intergovernmental panel on climate change
IQSY	International quiet sun year
LIS	Local interstellar spectrum
MCNP	Monte Carlo N-Particle transport code

NASA	National aeronautics and space administration
NGRIP	North GRIP
PCA	Polar cap absorption or principal component analysis
SEP	Solar energetic particle (event)
SCR	Solar cosmic rays
SFU	Solar flux unit
SN	Supernova
SPE	Solar proton event (archaic)
SPECMAP	Mapping spectral variability in global climate project
STE	Stratosphere–troposphere exchange
THC	Thermohaline circulation
TSI	Total solar irradiance
UV	Ultraviolet
VADM	Virtual axial dipole moment
YD	Younger Dryas

Part I

Introduction

Chapter 1

Motivation

For a small child, the world is full of wonders. Everyday it discovers new ones. Soon it starts to ask tough questions like: “Why does an apple fall down from a tree, why does the water in an alpine river with large rocks partly flow uphill, why is the Sun hot and why does it turn red in the evening, how is a rock formed and when did it happen”?

These and many other questions have been asked since ancient times and the more people thought about them and made experiments, the more they learned about our environment and the way nature works. Many basic laws of nature have been discovered during the past centuries. These laws seem to be universally valid for all times. They are quite different from the laws made by humans which are influenced by the cultural and historical background of people and are subject to frequent changes. Nature’s laws are the same in Europe, in Australia and on Mars. These laws help us understand what goes on, for example, deep in the Sun. By analyzing all the information available from the Sun in the form of light, particles (solar wind, neutrinos), magnetic fields, and the dynamics (sunspots, helioseismology) solar physicists are able to develop models of the interior of the Sun and how energy is generated there and transported out into the solar system. However, in spite of the impressive progress in natural sciences, with every answered question at least one new one arises and we are still far from understanding how nature really works.

We humans are living on the surface of planet Earth, two-third of which is covered by water, the essence of life. This water makes Earth look blue from space except for the continents and the white clouds which cover a considerable part as well. A very thin layer of gas, the atmosphere, separates us from space and protects us from harmful particles (cosmic rays) and radiation (UV, X-rays). All the men on the moon were deeply impressed when they saw the Earth for the first time from far away (Fig. 1-1).

The Earth is a very dynamic and complex system powered by solar energy. Life, which has existed for almost as long as the solar system, has modified the environment on Earth. Vegetation affects the hydrological cycle and changes the albedo,



Fig. 1-1 Earthrise seen from the moon. Apollo 8 crew Christmas 1968 (NASA)



Fig. 1-2 Earth seen from the Cassini spacecraft from behind Saturn (Cassini Imaging Team, SSL, JPL, ESA, NASA)

the portion of solar radiation which is reflected back into space. Photosynthesis alters the composition of the atmosphere, which was originally free of oxygen, by turning CO_2 into O_2 . In recent times, the growth of the Earth's population and the increasing consumption of energy per capita marks the beginning of a new era. By burning huge amounts of fossil fuels, we humans are increasing the atmospheric

CO₂ concentration to levels far above the values attained during at least the past million years. In other words, mankind has started a big experiment by considerably modifying the natural global conditions. Now, understanding how nature works and studying the consequences of our activities will be no longer simply an intellectual game for unworldly scientists. It is becoming a serious challenge with potentially large impacts on all of us living on the tiny spot called Earth photographed by the Cassini spacecraft from behind Saturn (Fig. 1-2).

To successfully address the question of global change in the past, present and future, we need to employ a holistic approach that crosses the boundaries between many disciplines. This book talks about cosmogenic radionuclides, radioactive isotopes which are produced and distributed within the earth system. It shows how they can be used to trace and to reconstruct the history of a large variety of processes. The authors hope to convey to the reader the beauty and the potential these tools have to contribute to the solution of many present and future problems.

Chapter 2

Goals

Richard Feynman (Feynman et al. 1963) began his famous “Lectures on Physics” with the question: *“If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words?”* His answer was: *“I believe it is the atomic hypothesis that all things are made up of atoms – little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.”*

We know today that everything in the universe, the stars, the planets and ourselves are made up of only about 90 different atoms (the elements) which themselves consist of just three particles: protons, neutrons, and electrons. A 100 years ago, only physicists and chemists studied and made use of atoms, and this was regarded then as “fundamental science”. The word physics, however, is derived from the Greek word *physis* (ΦΥΣΙΣ) and means “nature”, or the science of matter in space and time. All the natural sciences (and applications thereof) ultimately deal with the interaction and the motion of atoms and frequently the best (and sometimes the only) way to study these interactions is to trace the individual atoms themselves.

Today there is hardly a scientific profession that does not use atomic or nuclear techniques on a day to day basis – in lasers, in specialized semiconductors, in studying the growth of plants, in magnetic resonance tomography, and so on. This book deals with yet another application: the ability to use isotopes, and in particular the radioisotopes, to permit the study of a great number of environmental processes.

Until recently, the precise measurement of small quantities of atoms was an extremely difficult task, and this impeded the development of the applications that have blossomed over the past two decades. Stable isotopes are traditionally measured by mass spectrometers and both mass resolution and precision have steadily improved, together with a reduction in sample size. Radioisotopes had been measured by radioactive decay counting; however, this is very inefficient for radionuclides with long half-lives. The invention of the accelerator mass spectrometry technique (AMS) in the late 1970s revolutionized the field, enabling the measurement of isotopic ratios as low as 10^{-14} and increasing sensitivity by

5–6 orders of magnitude. The importance of such an increase may be illustrated by an analogy from astronomy. Imagine a Greek astronomer who was observing the stars with the 5 mm wide iris of his naked eye. Switching to AMS was equivalent to giving our ancient astronomer access to the 5 m telescope at Mount Palomar.

The cosmogenic radionuclides have been generated here on Earth, in quite miniscule amounts, from time immemorial. They were created by the cosmic radiation that bombards the Earth, and they provide a record of astronomical, solar, and Earth bound events that extend far into the past. Until recently, we did not have the means to read or decipher these records. However, as a result of the analytical advances outlined above, the number of applications of cosmogenic radionuclides to understand the past and investigate the present is growing rapidly and spreading over many scientific disciplines. The number of students and scientists working with cosmogenic radionuclides is steadily increasing, and with it the need for a comprehensive overview of what cosmogenic radionuclides are, and a discussion of their potential and their (present-day) limitations.

To date, some specialised review articles, conference proceedings, and a few books have appeared dealing with cosmogenic radionuclides as tracers and as dating tools, the primary focus being on specific fields of applications with only passing discussions of the basic underlying concepts and mechanisms. The main goal of this book is to provide the reader with a comprehensive discussion of the basic principles lying behind the applications of the cosmogenic (and other) radionuclides.

Part I (Chaps. 1–3) provides an introduction to the book and outlines the goals.

Part II (Chaps. 4–8) provides the background knowledge of the properties of the cosmic radiation that will allow the reader to understand the concepts, terminology, and formulae that are used later in the book.

Part III (Chaps. 9–15) of the book is dedicated to the cosmogenic radionuclides and discusses in some detail their production by the cosmic radiation, their transport and distribution in the atmosphere and the hydrosphere, their storage in natural archives, and how they are measured. A good understanding of these basics is a prerequisite for optimal use of cosmogenic radionuclides as environmental tracers and dating tools.

Part IV (Chaps. 16–23) deals with applications of cosmogenic radionuclides. It presents a number of examples selected to illustrate typical tracer and dating applications in a number of different spheres (atmosphere, hydrosphere, geosphere, biosphere, solar physics and astronomy). The goal of this part is not to give a comprehensive overview of all the many different applications developed so far. Its aim is to give the reader an understanding of what is possible, and possibly to provide the insight and motivation to develop new applications. At the same time, we have outlined the limitations of the use of cosmogenic radionuclides, to prevent unfortunate experiences in the future.

We are aware that we are addressing a wide audience, ranging from archaeology, biophysics, and geophysics, to atmospheric physics, hydrology, astrophysics and space science. We have therefore tried to explain everything at the level of a graduate student without specialist skills in physics or mathematics. To provide

the reader with more technical details, or the derivation of a few interesting equations, we use boxes which are separated from the main text and are not mandatory in order to understand the basic ideas. We would rather bore the specialists in their own field with “trivial” matter, than to keep others from understanding important concepts they are not familiar with. For the same reason, and to make reading individual chapters easier, we provide extensive internal referencing, and occasionally some basic information is repeated briefly in the introductions to chapters on specific applications.

Finally we express our hope that this book will help the reader to achieve a greater understanding of the use of the cosmogenic radionuclides, and also to enjoy the thrill of using this marvellous technology to uncover the mysteries of the past, and the world around us as we authors do.

Reference

Feynman RP, Leighton RB, Sands M (1963) The Feynman lectures on physics. Addison-Wesley, Menlo Park, CA

Chapter 3

Setting the Stage and Outline

In this chapter we set the stage for the rest of the book by introducing the terrestrial and space environments and the cosmogenic radionuclides. The main purpose of this general overview is to show that the terrestrial and space environments are large complex coupled systems and that cosmogenic radionuclides are extremely useful tools to better understand how these systems work and interact, and to reconstruct their history and even to some extent predict their future.

We stress the fact that the environment is a single, internally coupled system. By convention, and for simplicity, we usually divide it into subsystems; however, the interaction of these subsystems must always be kept in mind. A further complication is that there are no sharp boundaries between the sub-systems themselves. Thus the boundaries between the geosphere and the hydrosphere are indistinct, as are those between the atmosphere and space. It is impossible to pinpoint the top of the Earth's atmosphere because the density of the atmosphere decreases exponentially and merges into the space environment over a considerable distance. Here on Earth, the properties of artesian water depend on the geological setting. It is important to bear in mind that the terrestrial and space environments, and the various subsystems, are open systems which means that there is continuous exchange of energy and matter between them.

The cosmogenic radionuclides are an important consequence of processes occurring in the galaxy, the Sun, and in the atmosphere. They are generated by the cosmic radiation (protons, He, and heavier nuclei and electrons) that were accelerated to very high energies in supernova explosions within our galaxy and which have then travelled at almost the speed of light to the vicinity of our Sun, entering a region called the heliosphere. Their intensity is modified as they pass through the heliosphere, and then on entering the atmosphere they have collided with the nucleus of an atom and have broken it up into the smaller, unstable cosmogenic radionuclides (radioactive nuclides produced by cosmic rays).

In Fig. 3-1 we summarize the essence of this book in one single diagram. It is a sketch of the terrestrial and space environments consisting of the many components such as the land, ocean, lakes, rivers, vegetation, clouds, volcanoes, the Sun, cosmic rays, and so on. Traditionally the environment is divided into systems or spheres,

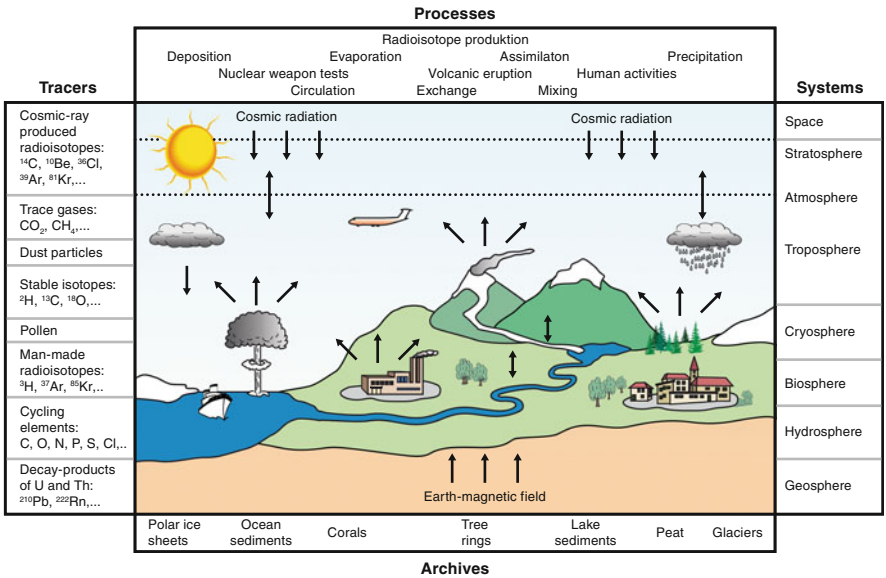


Fig. 3-1 The terrestrial and space environment divided into the subsystems listed on the *right-hand side*. Basic processes are listed in the *top panel*. They can be studied using tracers mentioned on the *left-hand side*. Information about the past environmental changes is stored in natural archives (*bottom panel*)

and these are indicated on the right-hand side of the figure. The environment is a dynamic system driven in large part by solar energy, in which many different processes are involved in the transport and distribution of energy and matter. Some of these processes are listed on the top of Fig. 3-1.

The basic goal of environmental science is to understand this large variety of physical, chemical, geological, and biological processes, their couplings and their spatial and temporal variability, and ultimately to construct an Earth system model which includes all the relevant processes. To achieve this goal we need to investigate the different processes themselves, and their interplay. However, it is often very difficult to study environmental processes because of their three-dimensional nature, their scale, and their inhomogeneity. A good example is the flow of underground water which is invisible and potentially very complex. We have long used “tracers” to this end; dyes and other chemicals that allow us to follow the motion of the water underground. The cosmogenic radionuclides play an important role here, as summarized in the list on the left-hand side of Fig. 3-1.

In general, tracers are tiny amounts of substances which are either naturally or artificially added to a system such as a water flow that allow us to document its movement over space and time. In the example of a groundwater system, the traditional approach was to use dyes and coloured material – we speak of such as having the property of “colour”. A tracer must fulfil several conditions: (1) it must be discernible from the medium it is added to, (2) it should be conservative – i.e. it

must behave in a predictable manner and not influence the system into which it is introduced, and (3) it must be easily detectable. As artificial tracers may cause disturbance of a natural system or social resistance (even if the amount is negligibly small) it is preferable to use natural tracers whenever it is possible.

The list of tracers in Fig. 3-1 shows that a considerable number of them are the naturally occurring cosmogenic radionuclides. The fact that they decay after a precisely known mean lifetime offers the additional opportunity to not only trace the pathway, but also to determine timing. In other words, while most tracers just carry “colour”, radionuclides carry a “colour” and a “clock” at the same time. Furthermore, due to their continuous production and radioactive decay, their source and sink functions are rather well known. The large range of different geochemical properties and half-lives of the cosmogenic radionuclides allows the user to select the one which is the most appropriate for the problem in hand.

The beauty of many tracers is that not only are they very useful tools in the study of processes going on at present, but they can also provide unique information about the past. This is particularly important because most of the important processes in the environmental system (such as climate change) occur on decadal to millennial or even geological time scales. While the amount of information provided by earth and space bound instruments has been almost unlimited for the past several decades, we face an increasing lack of environmental information the further we go back in time. Fortunately mother Earth has been keeping track of past environmental conditions in her own diaries, the natural archives. Many processes and parameters are recorded in ice cores, sediments, tree rings, and the other archives listed in the row at the bottom of Fig. 3-1. The cosmogenic radionuclides play a fundamental role here in establishing the time scales and in dating important past events. The information in the archives provides us a better understanding of the past, which is the key to predicting the future.

Having set the stage we now take the reader on a tour through the book. It is divided into four parts, starting with this introductory part. Part II provides an overview of the cosmic radiation. Chapter 5 starts with some historical remarks and definitions, and then outlines the origin and the properties of the galactic cosmic rays. The Sun’s effects on the propagation of cosmic rays through the heliosphere are summarized and then finally the effects of the geomagnetic field are outlined. Chapter 6 is dedicated to the instruments that have been used to measure the cosmic radiation. Chapter 7 discusses the various effects that lead to temporal variations of the cosmic radiation that, in turn, lead to temporal variations in the cosmogenic archives. Chapter 8 deals with the cosmic radiation produced by the Sun and its specific properties.

In Chap. 9, we go to Part III and introduce the cosmogenic radionuclides which are the product of the interaction of the cosmic radiation with matter. In Chap. 10, we address the interactions of cosmic rays with the atmosphere and how a cascade of secondary particles develops which finally leads to the production of the individual cosmogenic radionuclides. This chapter outlines how Monte Carlo techniques are used to compute the neutron and proton fluxes throughout the atmosphere (including the solar and geomagnetic effects) and then compute the cosmogenic production

rates. Chapter 11 discusses cosmogenic production in environmental systems other than the atmosphere, while in Chap. 12 other production mechanisms are addressed.

Chapter 13 deals with the vital matter of the manner in which atmospheric processes distribute and deposit the cosmogenic radionuclides on the Earth's surface. The geochemical cycles are different for the various cosmogenic radionuclides, as addressed in this chapter.

Some of the cosmogenic radionuclides are then incorporated into natural archives where their unique record of the past is stored over long periods of time. The most common natural archives and their main properties are described in Chap. 14. The last chapter of this part of the book deals with the techniques used to measure the cosmogenic radionuclides in natural archives.

The last part of the book is dedicated to applications. It presents a selection of typical examples that demonstrate the power of cosmogenic radionuclides to investigate various processes within the environmental system. The examples are from Solar Physics and Astronomy in Chaps. 17 and 18 and from the other systems, shown in Fig. 3-1: the atmosphere (19), the hydrosphere (20), the geosphere (21), and the biosphere (22). The last chapter of this book finally discusses the applications related to dating past events.

Part II

Cosmic Radiation

Chapter 4

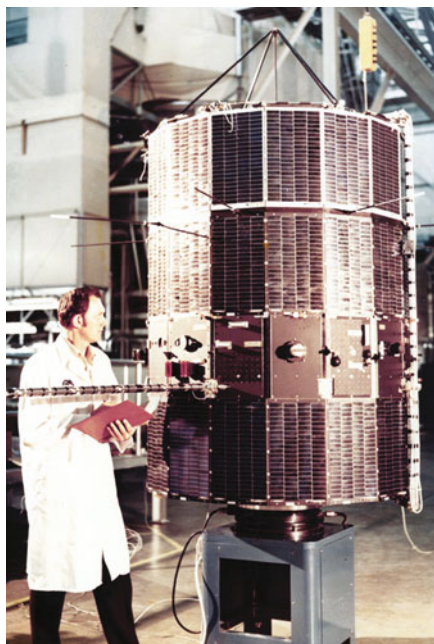
Introduction to Cosmic Radiation

Long before the first animals emerged from the oceans to live on land, before the rise and fall of the dinosaurs, and long before the Neanderthals roamed the Earth, supernovae throughout our galaxy were happily making cosmic rays that then bombarded Earth. Long before mankind developed the ability to keep records, the cosmic rays were leaving behind indelible records that now, millions of years later, we have learned to read. Those records were written in an ink that consisted of the “cosmogenic radionuclides” produced by the cosmic rays when they smashed asunder atoms in our atmosphere. Those cosmogenic records tell us about the past behaviour of our Earth, our Sun, and our galaxy. They provide us with the means to study our atmosphere, water resources, to date archaeological discoveries, and do many other things that enrich our lives.

The cosmic radiation has complex characteristics, and varies from day to day, year to year, century to century and over timescales of many thousands of years. These characteristics, and time variations, both help and hinder our interpretation of the cosmogenic records. They allow us to use cosmogenic data to investigate the manner in which the Sun has varied over the past millennia that will assist us to understand climate change, and to avoid expensive damage to our satellite communication systems. At the same time, they complicate the carbon dating of archaeological materials. To exploit the good things, and minimise the bad, it is necessary to understand (or at least, look up) the properties of the cosmic rays without which we would have no cosmogenic records, or practical applications.

It is only in the past two decades or so that we have had the means to determine, in a quantitative manner, the relationship between the cosmic radiation and the production of the cosmogenic radionuclides. Two major technological advances made that possible. (1) Instrumentation on earth orbiting satellites and deep space probes provided a detailed knowledge of the composition, energy dependence, and time dependence of the cosmic radiation. (2) Large numerically intensive computers and highly complex mathematical codes that simulate the propagation of the cosmic rays in the atmosphere provided the ability to use satellite data to interpret the cosmogenic data from the past. The development of the Accelerator

Fig. 4-1 The sixth Interplanetary Monitoring Platform (IMP 6) during pre-flight “check-out”. The electrical power to operate the scientific systems and the data transmitter were supplied by the solar cells that cover the top and bottom sections of the cylindrical body. The cosmic ray and other instruments peer out into space from the cylindrical section of the spacecraft adjacent to the engineers’ head (NASA)



Mass Spectrometer (AMS) at about the same time provided another and major stimulus to the use of cosmogenic radionuclides in a wide range of applications.

Figure 4-1 is an image of the sixth Interplanetary Monitoring Probe (IMP 6) that contributed greatly to our present-day ability to interpret the messages conveyed by the cosmogenic radionuclides. It was launched into a highly eccentric orbit around the Earth in 1971, and it and similar spacecraft (IMP 7 and 8) carried cosmic ray detectors that provided a continuous record from 1973 to 2006. In addition, it carried instruments to measure the strength of both the heliospheric magnetic field and the solar wind, and all of these contributed greatly to our present day understanding of the influence of the Sun on the production of cosmogenic radionuclides here on Earth. It is this understanding that allows us to interpret the cosmogenic data from the far past in terms of the highly detailed knowledge of the Sun and Earth that we have acquired during the space age.

The following four chapters regarding the cosmic radiation (Chaps. 5–8) provide the historical, instrumental, and theoretical background to the later chapters in this book. We also hope that they may be useful as a “stand-alone” reference for a wider audience that has a need for an up-to-date account of the more important facets of the cosmic radiation.

Chapter 5

The Cosmic Radiation Near Earth

5.1 Introduction and History of Cosmic Ray Research

The half-lives of the cosmogenic radionuclides are very short compared to the age of the Earth, and none would now exist on Earth without their continuous production through the interaction of cosmic rays with the atmosphere. Seventy-five years of instrumental observations have shown that the intensity of ~ 3 GeV cosmic rays changes by $\geq 20\%$ over time scales of ~ 10 years, superimposed upon longer term changes that are not well defined in the instrumental record (see Chaps. 6 and 7). The geomagnetic field has a strong screening effect upon the cosmic radiation as well and, as a consequence the production rates of the cosmogenic radionuclides near the equator are approximately 10% of those in the polar caps. This screening effect is strongly influenced by the strength and configuration of the geomagnetic field, which also changes greatly over time. Any study of the cosmogenic radionuclides, or their utilization for scientific or practical purposes therefore demands an understanding of the properties of the cosmic radiation itself and the screening imposed by the geomagnetic field.

This chapter outlines the origin and properties of the “galactic” cosmic radiation (GCR) and describes the manner in which the Sun influences the intensity in the vicinity of Earth. It then outlines the screening effects of the geomagnetic field and describes the “modulation function” that is widely used in cosmic ray and cosmogenic studies and in a number of practical applications such as space weather.

A hundred years ago, the scientific world was agog with excitement. Radioactivity had been discovered. The radioactive elements were found to emit radiation that passed through metals and soon, the alpha (helium nuclei), beta (electrons) and gamma (electromagnetic quanta) radiations had been discovered. Hertz had demonstrated the existence of radio waves. Roentgen discovered that X-rays could be generated in an electrical discharge. All forms of radiation were investigated with great vigour.

The “electroscope” was a commonly used radiation detector at that time. It used electrostatic forces to measure the amount of electrical charge on an electrical

capacitor, using air as the dielectric between the plates. The presence of radioactive radiation caused the charge to leak away due to ionization of the air in the capacitor. The stronger the radiation, the faster the charge leaked away. Over time, careful experiments showed that there was another radiation, much more penetrating than the alpha, beta, and gamma radiation already discovered in the laboratory.

Early on the morning of 7 August, 1912, the Austrian physicist Viktor Hess and two companions commenced a balloon flight that reached an altitude of ~5 km before landing near Berlin. Hess observed that the discharge rate of two of his electrometers initially decreased up to an altitude of about 1,500 m, and then began to rise, until at 5,000 m it was almost twice the rate at the surface. He published his results in November of that year, writing; “The results of these observations seem best explained by a radiation of great penetrating power entering the atmosphere from above.” This radiation was initially called “*Ultrahochenergiestrahlung*” – “ultra-high energy radiation.” Its apparent origin in the cosmos led to it becoming known as “the cosmic radiation.” Twenty-four years later, Hess shared the Nobel Prize in physics for his discovery of cosmic radiation.

For 20 years, cosmic rays presented a major scientific enigma. What were they? Where did they come from? The Nobel Laureate R. A. Millikan argued strongly that they were very energetic gamma rays; he called them the birth cries of the atoms, until it was shown in the early 1930s that they were deflected by the Earth’s magnetic field. This proved that they were electrically charged particles, and this initiated the tortuous process that has lead to today’s understanding of these extremely energetic protons and other elemental particles of matter.

Starting in the 1930s, there have been two primary sources of information about cosmic radiation. Continuously recording instruments on the surface of the Earth (ionization chambers, Geiger counter telescopes, and neutron monitors, (see Chaps. 6 and 7) have shown that the intensity of the radiation varies with time, energy, and direction in space. Measurements from balloons, and more recently satellites, have allowed us to examine the radiation before it interacts with atmospheric nuclei, permitting us to determine its elemental and isotopic composition and to extend our knowledge to the low energy particles that cannot penetrate even to balloon altitudes (<100 MeV/nucleon). These features of the cosmic radiation are summarized in the following sections, since all of them influence the production of cosmogenic radionuclides in the atmosphere and the space weather effects near Earth.

Because of these influences, the cosmogenic data themselves have become our third means to study cosmic radiation. Unlike the surface and satellite detectors which measure the instantaneous properties of the radiation, cosmogenic data provide information on how the cosmic radiation has varied in the past. In this role, we refer to them as providing the “paleocosmic radiation record.” As discussed in the following chapters, the modern measurements made by surface and satellite-borne cosmic ray detectors provide us with a “Rosetta Stone” that allows us to decipher the cosmogenic data to understand the properties of the cosmic radiation, and the solar system, over the past millennia.

5.2 The “Rosetta Stone” of Paleocosmic Ray Studies

It was realized from the beginning that the cosmogenic radionuclides would be produced by the galactic cosmic rays [e.g., (Lal and Peters 1962)] and, over time (by about 1980) it was agreed by some investigators that the temporal variations of ^{14}C over the past 1,000 years, say, were correlated with solar activity [e.g., (Eddy 1977; Stuiver and Quay 1980)]. This was not a universal view: some in the cosmic ray community argued that there were decadal and longer scale variations in the cosmogenic data that were not seen in the instrumental record, and that this must indicate that there was some major, non-cosmic ray influence that was the dominant factor. Others in the cosmogenic community argued that the long-term variations in the ^{10}Be data might be due to changes in the Earth’s climate. Furthermore, there was no quantitative means to relate the variations in cosmogenic data to changes in the intensity of the cosmic radiation outside the magnetic field of the Earth. Without that, there could be no understanding of how the cosmogenic record related to the cosmic ray measurements of the instrumental and space eras. The situation was similar to that in archaeology where no one knew how to translate the hieroglyphic records of the Egyptian civilization until the “Rosetta Stone” was discovered in 1799, in which the same government announcement was given in hieroglyphics and in Greek. The information on that tablet provided the means to decode the hieroglyphics. Studies in nuclear physics and of the instrumental cosmic ray records (ground level and space) in the 1970–1990s provided the equivalent of the Rosetta stone that has allowed the cosmogenic record to be deciphered.

The most important component of the paleocosmic ray Rosetta Stone was provided by the development of mathematical codes by the nuclear research community to simulate the passage of energetic charged particles through matter (see Chap. 10). First used to model the production of cosmogenic radionuclides in meteoritic matter (Masarik and Reedy 1995), they were soon used to compute the counting rates of neutron monitors (Chap. 6) and the production of the cosmogenic radionuclides in the atmosphere (Chap. 10). Similar codes were developed to compute the effects of the intense solar cosmic radiation upon the polar atmosphere (Chap. 8). The instrumental and space measurements allowed the energy dependence of the cosmic ray spectrum in interstellar space to be estimated and the manner in which solar activity influenced the spectrum at Earth. Taken together, this information demonstrated that the cosmogenic data were consistent with the measurements of the instrumental and space eras and allowed quantitative intercalibration of the paleocosmic radiation to the present. As with the original Rosetta Stone, this then allowed other factors to be deciphered. One of the most important (and quite recent) additions was the use of modern global climate models to compute the manner in which the cosmogenic radionuclides are mixed in the atmosphere prior to sequestration in polar ice (Chap. 13).

Throughout this book, we will refer to the key role that the paleocosmic ray Rosetta Stone has played in the past 20 years. Without those advances, the greater

majority of the material in this book and in the modern paleoc cosmic ray and cosmogenic nuclide literature would not exist.

5.3 Some Important Definitions

All fields of science develop their own systems of specialist units that may be very confusing to the “outsider.” This is particularly so in a book such as this which sits at the interface between a number of specialist areas. With this in mind, we have adopted the following policy (a) to use mks units for general discussions and (b) to use the specialist units from glaciology, astronomy, cosmic ray physics, etc, wherever appropriate. When specialist units are first introduced, we have defined them in terms of the mks system and also have provided a glossary at the back of the book.

There are a number of concepts that are central to understanding the contents of this book, which we summarize in the following. It is important to note that the majority of the cosmic radiation is travelling at speeds close to that of light, and consequently it is mandatory that we use “relativistic” expressions throughout. We stress, however, that we have written the book on the basis that the majority of readers will have no need to understand the complexities of relativistic physics. All they need to know is that those complexities exist and have been taken into account throughout the book. For the reader seeking a brief introduction to the equations and concepts of the “special theory of relativity,” Box 5.3.1 outlines those concepts that are relevant to the study of the cosmic radiation and the cosmogenic radionuclides.

5.3.1 BOX Relativistic Physics

In 1905, Albert Einstein stated his theory of “special relativity,” that explains how the concepts of “classical” (or “Newtonian”) physics change as objects move at speeds approaching that of light. This is true of the majority of the cosmic radiation, and so his equations must be used in their study. For the purposes of this book, it is sufficient to outline three of the key concepts and the equations that describe them.

- Relativistic mass
- The “dilation” of time
- Energy and momentum at relativistic speeds

Experiments in the late nineteenth century led to the conclusion that neither matter nor waves could move with a speed greater than the speed of light. This led to the concept of the “relativistic mass” of a rapidly moving particle:

$$m = \frac{m_0}{\sqrt{1 - (v/c)^2}} \quad (\text{B5.3.1-1})$$

where m_0 is the “rest mass”, and v and c are the speeds of the particle and light. The equation states that the apparent mass (to satisfy Newton’s second law of motion) of the particle increases rapidly for large velocities ($>0.1c$, say).

Einstein’s theory also led the equation for the “dilation of time”:

$$\Delta t_0 = \frac{\Delta t_p}{\sqrt{1 - (v/c)^2}} \quad (\text{B5.3.1-2})$$

where Δt_0 and Δt_p are the durations of some event, as seen by an observer on Earth, and by the moving particle itself. Thus, the equation says that the event is seen to take longer by the observer on Earth than by a very fast moving particle. A good example is the case of the muon, as discussed in Sect. 10.2.1. The fast muons produced by incoming cosmic rays have a mean lifetime of 2.2 μs , and since they are moving with a speed close to that of light ($3 \times 10^8 \text{ ms}^{-1}$), they travel on average 660 m before they decay. Yet many reach the surface of Earth 20 km below. The reason is that the time dilation equation tells us that the half-life, as seen by an observer on Earth, is greater than that “seen” by the muon itself. For example, muons travelling at speeds of $0.9c$ and $0.99c$ would be seen from Earth to have half-lives of 6.9 and 21.3 μs , respectively. A companion Einstein equation tells us that for the muon, the distance to the surface of Earth is reduced by the same factors, so there is no conflict between analysing the process in terms of us on Earth, or the muon itself.

Finally, we outline the mathematics that led to the relationships summarized under “Energy” and “Rigidity and Gyroradius” in Sect. 5.3.

The momentum p is given from another of Einstein’s equations by

$$p^2 c^2 = T^2 - m_0^2 c^4 \quad (\text{B5.3.1-3})$$

This gives pc in units of eV (or MeV), the same units as used for the total energy T (rest plus kinetic) and $m_0 c^2$, the rest energy.

For galactic cosmic rays (which are totally ionized) the rigidity, P , is then defined as

$$P = pc/q = pc/Ze \quad (\text{B5.3.1-4})$$

where Z is the atomic number. If pc is in eV (or MeV), then pc/e is in volts, V (or MV).

Now for an ion gyrating with velocity $\mathbf{v}$ perpendicular to $\mathbf{B}$, the force on it is $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$ and it is balanced by the centrifugal acceleration:

$$m\mathbf{v}_\perp^2/\rho_{\text{gyro}} = q\mathbf{v}_\perp \mathbf{B} \quad (\text{B5.3.1-5})$$

(continued)

So

$$\rho_{\text{gyro}} = mv/q\mathbf{B} = mvc/q\mathbf{B}c = P/\mathbf{B}c \quad (\text{B5.3.1-6})$$

The Differential Cosmic Ray Intensity. This specifies the manner in which the cosmic radiation intensity varies with (usually) kinetic energy, E , direction in space, $\mathbf{D}$, and time, t . Let dN be the number of cosmic ray protons in the energy interval, E to $E+dE$, passing through a surface of area S within time dt , from within a solid angle of $d\Omega$, from direction $\mathbf{D}$. Then the proton differential intensity at energy E , from the direction $\mathbf{D}$, is defined as

$$j_p(E, \mathbf{D}, t) = dN/(S \cdot dt \cdot dE \cdot d\Omega) \quad (5.3-1)$$

and is usually measured in either protons per square meter-steradian-sec-GeV or protons per square meter-steradian-sec-MeV. In the literature, the differential spectrum will be found defined by either $j()$, or $J()$.

The cosmic radiation consists of protons, helium, heavier nuclei, and electrons and the differential intensity of each may be calculated using (5.3-1).

In our part of the galaxy, it has been found that for $E > 20$ GeV, to first approximation, $j_p(E)$ can be expressed as a power law $j_0 E^{-\gamma}$, where γ is called the *spectral exponent*, and j_0 sets the absolute magnitude. This will be discussed further in the next section.

The Integral Cosmic Ray Intensity. The proton “integral intensity” for energy E^* is the total flux of protons above energy E^*

$$\begin{aligned} J_p(E^*, \mathbf{D}, t) &= dN/(dS \cdot dt \cdot d\Omega) \\ &= \int_{E^*}^{\infty} j_p(E, \mathbf{D}, t) dE \end{aligned} \quad (5.3-2)$$

and is quoted in the units of protons per square meter-steradian-second. The integral spectra of He and heavier cosmic rays are defined in a similar manner. As a result of the integration with respect to energy in (5.3-2), the spectral exponent of an integral spectrum is 1 less than the exponent of the corresponding differential spectrum. Thus for a differential spectrum with $\gamma = 2.37$, the spectral exponent for the equivalent integral spectrum equals 1.37.

Units of Energy. The energy of a cosmic ray particle is usually stated in electron volts (eV). This is the energy gained when a particle bearing a single electronic charge crosses a potential difference of 1 V. Thus, both a proton and an electron gain 1 eV of energy when accelerated by 1 V. One eV equals 1.6×10^{-19} J. In the majority of applications in this book and in the literature, MeV (10^6 eV), GeV (10^9 eV), and occasionally, TeV (10^{12} eV) are used.

The total energy, T (rest energy plus kinetic energy) of a cosmic ray of mass m_0 is given by

$$T = \frac{m_0 c^2}{\sqrt{1 - (v/c)^2}} = m_0 c^2 + E = E_0 + E \quad (5.3-3)$$

where v and c are the speed of the cosmic ray and light, respectively. The “rest mass” energy of the proton is $m_0 c^2 = 938 \text{ MeV}$, and E is the kinetic energy.

A significant fraction of the cosmic radiation consists of nuclei heavier than the proton. In such cases, the energy of the cosmic ray is usually quoted in terms of the energy per nucleon. Thus, a ${}^4\text{He}$ cosmic ray has a kinetic energy of $4E$, where E is the kinetic energy in eV/nucleon.

Rigidity and Gyroradius. The majority of cosmic rays carry an electrical charge, and they are therefore deflected by magnetic fields. The vector force, $\mathbf{F}$ on a cosmic ray particle is given by the vector product

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \quad (5.3-4)$$

where $\mathbf{v}$ is the velocity vector of the cosmic ray, $\mathbf{B}$ is the magnetic induction vector, and q is the charge. If the angle between $\mathbf{v}$ and $\mathbf{B}$ is θ , the magnitude of this force is $q\mathbf{v}\mathbf{B} \sin(\theta)$. This force causes the cosmic ray to follow a curved path, where the radius of curvature (the “gyroradius”) is given by

$$\rho_g = p_{\text{perp}}/q\mathbf{B} \quad (5.3-5)$$

where p_{perp} is the component of momentum perpendicular to the magnetic vector. We define a new quantity, the particle rigidity

$$P = pc/q$$

which is measured in units of volts, or more commonly for cosmic rays, MV (10^6) or GV (10^9).

Then the gyroradius of the particle is given by

$$\rho_g = P_{\text{perp}}/Bc \quad (5.3-6)$$

In mks units, we express P in volts (V), $\mathbf{B}$ in Tesla, and c in m/s; so ρ is in meters. Substituting for c and changing V to GV yields the useful relationship

$$\rho_{\text{gyro}}(m) = 3.3P(\text{GV})/B(T) \quad (5.3-6a)$$

where the appropriate units are given in parenthesis. Thus a 1 GV cosmic ray in a 1 nT field has a gyroradius of $3.3 \times 10^9 \text{ m}$. Interplanetary and interstellar magnetic fields are frequently quoted in the specialist unit, gamma = $10^{-5} \text{ G} = 10^{-9} \text{ T} = 1 \text{ nT}$.

Relationships between rigidity, P , and energy per nucleon, E , are given by

$$P = \frac{A}{Z} \sqrt{(E_0 + E)^2 - (E_0)^2} \quad (5.3-7a)$$

$$E = \left\{ \sqrt{\left(\frac{ZP}{A}\right)^2 + (E_0)^2} \right\} - E_0 \quad (5.3-7b)$$

where E_0 is the rest energy of a nucleon = 938 MeV and Z and A are the atomic number and atomic mass numbers of the cosmic ray particle, respectively. That is, the cosmic ray carries Z electronic charges and A nucleons (protons plus neutrons). For highly relativistic cosmic rays ($E \gg E_0$), $P = (A/Z)(E + E_0)$, while for non-relativistic particles the rigidity varies as the square root of E . Figure 5.3-1 plots rigidity (P) against kinetic energy per nucleon (E) for protons, helium, and heavy cosmic rays. For interest, the curve for an electron is given as well.

It is important to note the factor (A/Z) in these equations. $A/Z = 1$ for a proton, equals 2 for ^4He , and approximates 2 for all heavier cosmic rays. As a consequence, alpha particle and “heavy” cosmic rays have twice the rigidity of a proton with the same energy per nucleon. That is, Eq. (5.3-6a) tells us that the gyroradii of alphas and heavies are twice those of protons, that is they are less deflected by magnetic fields than the protons. The motion of cosmic rays in space and in the geomagnetic field is strongly influenced by their gyroradii, and consequently the alphas and heavies behave differently than protons in a number of important ways that are of significance in analysis of the cosmogenic radionuclides.

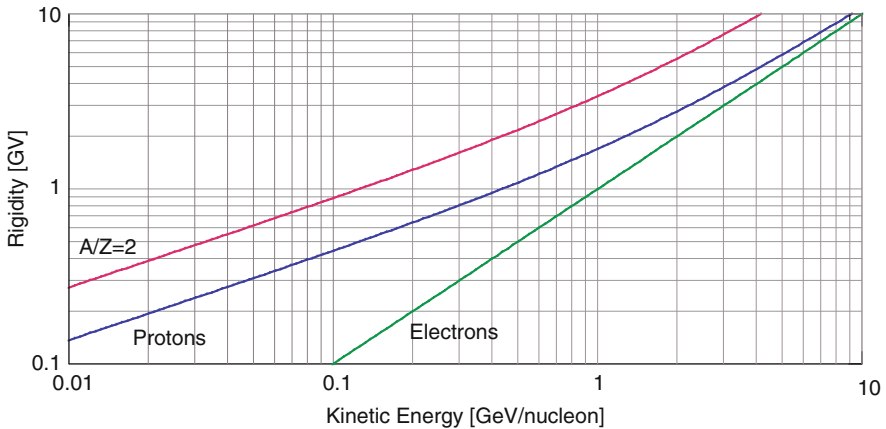


Fig. 5.3-1 The relationship between rigidity and energy per nucleon as a function of A/Z . The upper line represents alpha particles and heavier cosmic rays with $A/Z = 2$, followed by protons and electrons. For $E > 10$ GeV/nucleon, P [GV] $\approx (A/Z)(E + 1)$ [GeV]

The Astronomical Unit, the Light Year, and the Parsec. An astronomical unit (1 AU) is the mean distance between Earth and Sun: = 149,597,871 km. A light year is the distance travelled by light in a year = 9.4607×10^{12} km. A parsec (pc) is defined as the distance from which one AU subtends an angle of one arc second at the observer. One pc = 30.857×10^{12} km = 206,260 AU = 3.2616 light years.

Isotropy and Anisotropy. The cosmic radiation is “isotropic” when the intensity is identical from all directions at the point of interest. When it is not the same from all directions, it is said to be anisotropic.

Half-life, Decay constant, and Mean lifetime. The decay of a radionuclide can be described equally well in terms of the half-life ($T_{1/2}$), the decay constant (λ), or the mean lifetime (τ):

$$T_{1/2} = \frac{\ln(2)}{\lambda} = \ln(2)\tau \quad (5.3-8)$$

5.4 The Origin and Properties of the Galactic Cosmic Radiation

Essentially, all of the high energy (>20 GeV) cosmic rays reaching the top of the atmosphere have come from our own galaxy. At lower energies, there are occasional impulsive bursts of cosmic rays generated near the Sun (see Chap. 8), and there is a flux of low energy (<50 MeV) “anomalous cosmic radiation” accelerated in the heliosphere. The galactic cosmic rays (GCRs) largely originate in supernovae which occur on average once every 30–50 years in our galaxy. A supernova consists of the cataclysmic collapse of a heavy, old star, in which a very large amount of gravitational energy ($\sim 10^{51}$ erg) is converted to kinetic energy. Figure 5.4-1 depicts the Crab supernova and shows the debris that was ejected from the supernova after it collapsed. Analysis of photographs such as this shows that the expansion velocities are very high, indicating that highly supersonic shock waves travelled through the interstellar medium (gas and magnetic field) that surrounded the star following the initial explosion (Schlickeiser 2003).

Theoretical studies, and observations in the Earth’s magnetosphere (Sect. 5.8.1), show that shock waves in a magnetized plasma can accelerate ions and electrons to very high energies. As a consequence, supernovae are very strong emitters of radio waves, indicating the presence of a very large population of fast electrons accelerated by the expanding shock waves. They continue to radiate radio waves for many millennia, and the radio observations of many supernovae show that the particle acceleration processes are highly efficient. Satellite and ground level gamma ray telescopes show that supernovae also emit high-energy gamma rays, consistent with the newly accelerated cosmic ray protons and heavier nuclei colliding with gaseous atoms in the ejecta from the original explosion. The strongly polarized nature of the light and radio waves shows that strong magnetic fields extend throughout the exploding material. These and other observations of a number of supernova in our own, and neighbouring galaxies, has allowed determination of the physical conditions in the supernova and their accompanying shock

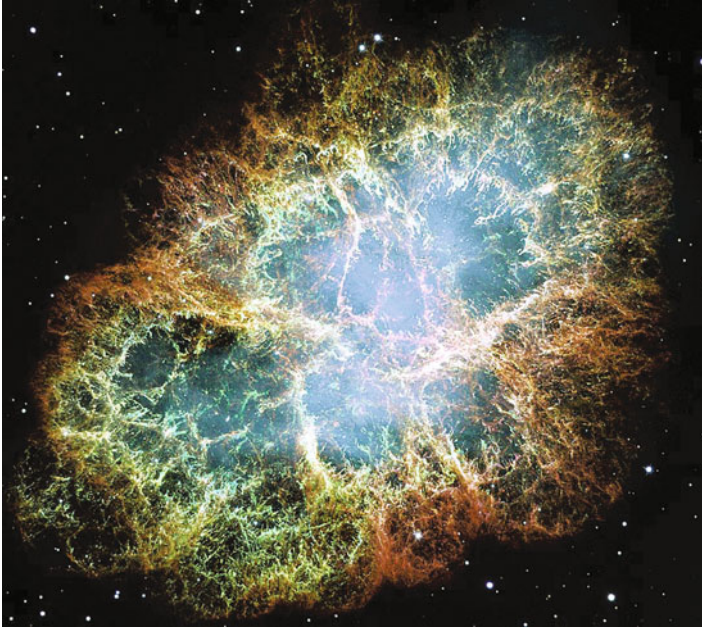


Fig. 5.4-1 The Crab Nebula, the remnants of a supernova observed by Chinese astronomers in 1054 AD. Although 2 kpc from Earth (4×10^8 times further from us than the Sun), it was bright enough to be seen in daylight when first seen (NASA)

waves. They lead to the conclusion that supernovae are prodigious sources of relativistic cosmic rays (≥ 1 GeV/nucleon), with some particles attaining extremely high energies ($> 10^6$ GeV/nucleon for ions).

Theoretical studies also show that highly supersonic shock waves will result in the differential energy spectra $j(E)$ of the electrons and ions varying as E^{-2} in all supernovae. Since supernovae occur in old stars, the gas in the vicinity of the star contains heavier elements (a substantial quantity of the hydrogen having been converted to heavy elements by nuclear fusion). The shock waves accelerate the hydrogen and heavier nuclei in a similar manner, and as a consequence, the isotopic composition of the ionic cosmic radiation is believed to be broadly similar in all supernovae.

Our galaxy contains $\sim 10^{11}$ stars, within a disc-shaped volume of radius $\sim 15,000$ pc, and ~ 500 pc thick ($1 \text{ pc} = 3.26 \text{ light years} = 3.086 \times 10^{16} \text{ m}$). It is believed to be similar to the Andromeda spiral galaxy depicted in Fig. 5.4-2. Our sun orbits the centre of the galaxy with a period of 200–250 million years. The stars in our galaxy are congregated along spiral arms that are a common feature of galaxies (see Fig. 18.2-1). The “interstellar space” between the stars in the galaxy contains a tenuous gas ($\sim 0.1 \text{ atoms cm}^{-3}$) and a seemingly weak magnetic field ($\sim 0.1 \text{ nT}$). Following their acceleration in a supernova, the charged cosmic ray ions and electrons escape into the local interstellar environment. However, they cannot



Fig. 5.4-2 The Andromeda spiral galaxy, M31, believed to be similar to our own galaxy (NASA)

travel very far. As described in Sect. 5.3, they are constrained to move in roughly circular orbits by the interstellar magnetic field, the gyroradius of a 100-GeV cosmic ray proton in a 0.1-nT field being only 10^{-4} pc. Irregularities in the interstellar magnetic field, and its overall alignment along the spiral arms, means that the cosmic rays “diffuse” slowly outwards away from their supernova of birth, and therefore their contribution to the cosmic ray intensity elsewhere in the galaxy changes very slowly with time. Because the galactic magnetic field inhibits them from escaping from the galaxy easily, the average age of a cosmic ray in the galaxy is 10^7 years (This number is known from a study of the elemental abundance of the cosmic rays themselves; see below). As a consequence, the cosmic ray intensity at any point in the galaxy is an average over contributions from a large number of individual supernova sources that have occurred in the distant past.

The spiral structure of the galaxy and other factors mean that the density of stars, their velocity of motion, the gas density, and the magnetic field all vary from point to point within the galaxy. As a consequence, the cosmic ray intensity is also expected to vary throughout the galaxy. The Earth is situated two-thirds of the radius from the centre of the galaxy, and recognizing the non-uniform nature of the galaxy, we speak of its “local interstellar environment.” In particular, the spectrum and intensity of the cosmic radiation in the vicinity of the solar system is called the “local interstellar spectrum” (LIS). The location of the Sun within the galaxy and the galactic arm has changed slowly with time, and it is reasonable to expect that there have been slow changes in the LIS as a consequence. For example, the Sun moves into and out of a spiral arm approximately every 70 million years, and on a shorter time scale, it entered a lower density region in the spiral arm of our galaxy $\sim 60,000$ years ago, and it is possible that the LIS changed at that time (see Sect. 18.2).

That is, the LIS may have changed over time for two different and independent reasons. In the first place, there may have been a slowly changing component due to charged cosmic rays slowly diffusing from a number of supernovae in the neighbourhood of our Sun. Secondly, our Sun's motion in the galaxy may have moved it into a region with a different LIS. In general, then, the local interstellar cosmic ray intensity near Earth is expected to have varied slowly with time. If this occurred, it will have influenced the production of the cosmogenic isotopes on Earth and in meteoritic matter in the solar system. Studies of the radioactive nuclides in meteorites and lunar material have indicated that the LIS has been constant to within a factor of 2 over the past 10^9 years. However, they also suggest that for the past 4×10^5 years, the LIS has been 50% higher than the 10^9 -year average (Chap. 18).

Against that background, we now discuss the present-day nature of the LIS near Earth. As discussed above, theoretical considerations indicate that many supernovae have injected cosmic rays with an E^{-2} spectrum into the galactic magnetic fields. The gyroradii of the cosmic rays increase almost linearly with energy, and so the higher energies are less constrained by the galactic magnetic fields. Over millions of years, some of them have escaped preferentially into intergalactic space. As a consequence, there has been selective loss of the higher energies during the 10^7 years that the average cosmic ray stays in the galaxy (see below). This leads to the steeper $E^{-2.37}$ spectrum shown in Fig. 5.4-3 for $E > 10^{11}$ eV (i.e., 100 GeV).

Examination of Fig. 5.4-3 shows that the cosmic ray spectrum exhibits an approximately power law spectrum over a remarkable ten orders of magnitude from 10^{11} to 10^{21} eV. While the cosmic ray intensity is 1 particle per square meter per second at 10^{11} eV (100 GeV), it is only 1 particle per square kilometre per year

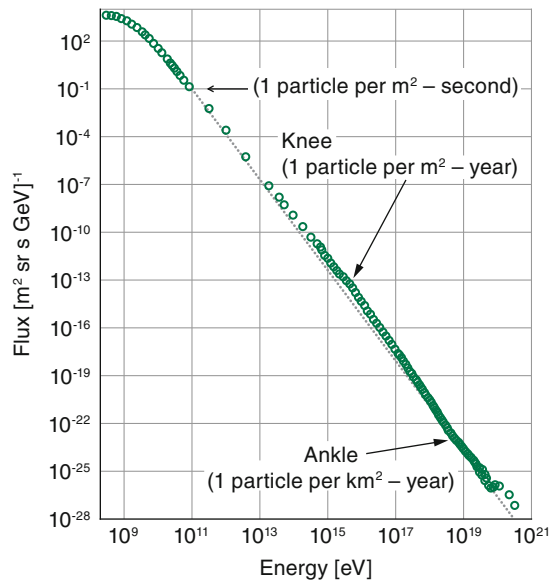


Fig. 5.4-3 The differential energy spectrum of the proton component of the cosmic radiation near Earth. The shape of the spectrum below 2×10^{10} eV (20 GeV) is determined by the level of solar activity (Chap. 7). In the case shown here, the Sun was relatively inactive. After Swordy (2001)

near 10^{19} eV (10^{10} GeV). In parenthesis, we note that these very infrequent, high-energy cosmic rays are detected by measuring the particles and electromagnetic quanta produced in the cascade initiated when the cosmic ray collides with an atom in the atmosphere (Chap. 10). The instrumentation for this is called an “air shower array;” however we will not discuss it any further in this book.

The spectrum in Fig. 5.4-3 has a distinctive change in slope (called the “knee”) at $\sim 10^7$ GeV, which has been attributed to less efficient trapping of the galactic magnetic fields for these higher energy cosmic rays. Another change in slope (the “ankle”) occurs near 10^{10} GeV. At energies < 20 GeV, the spectrum is affected by solar effects and the shape of the spectrum, and the intensity varies with time (see Sect. 5.7 and Chap. 7).

While the energy spectrum in Fig. 5.4-3 extends to very high energies, the fluxes are small, and as a result the preponderance ($> 99\%$) of the cosmic radiation reaching Earth has an energy < 100 GeV. In a similar manner, the proton and helium cosmic rays are the main contributors to the instrumental and cosmogenic measurements of the cosmic ray flux. As will be discussed in Sect. 5.7, the intensity of the lower energy cosmic radiation is strongly influenced by the degree of solar activity, leading to $\sim 50\%$ variations in the production rates of the cosmogenic radionuclides on Earth.

Following slow diffusion from many supernovae, the cosmic radiation in our part of the galaxy is essentially isotropic and comprises electrons, positrons, and atomic nuclei. The composition of the latter is compared with that of the matter in the solar system and the local galaxy in Fig. 5.4-4. Clearly, H and He are the most common nuclei: $\sim 87\%$ of the cosmic rays being protons and $\sim 12\%$ He and all heavier nuclei totalling $\sim 1\%$. These ratios depend upon energy and on the degree of solar modulation, and $\pm 1\%$ and greater differences in the H and He components will be found in the literature. As Fig. 5.4-4 shows, the light nuclei Li, Be, and B are 5–6 orders of magnitude more abundant in the cosmic radiation than in the solar system and the local galaxy; likewise, the nuclei between Sc and Mn are approximately 2–3 orders of magnitude high. These greatly enhanced contributions are attributed to the collision of the heavier cosmic rays (e.g., Fe) with interstellar gas, the resulting spallation reactions yielding a wide range of nuclei of lower atomic number. These high abundances of Li, Be, and B, and the presence of relatively short lived radioactive nuclei (e.g., ^{10}Be and ^{14}C) in the incoming cosmic radiation have led to the conclusion that the average lifetime of the average galactic cosmic ray is 10^7 years and that, on average, the cosmic rays in the vicinity of Earth have passed through $6\text{--}9 \text{ g cm}^{-2}$ of interstellar matter since their birth.

It is important to note that none of the He or heavier nuclides in the galactic cosmic radiation survive entry into the Earth’s atmosphere –all disintegrate into protons and neutrons, and together with the proton component of the galactic cosmic radiation, initiate the nucleonic chain that generates all of the cosmogenic radionuclides we observe on Earth (Chap. 10). In particular, while there are ^{10}Be and ^{14}C cosmic rays, they do not survive to become part of the cosmogenic radionuclides observed on Earth.

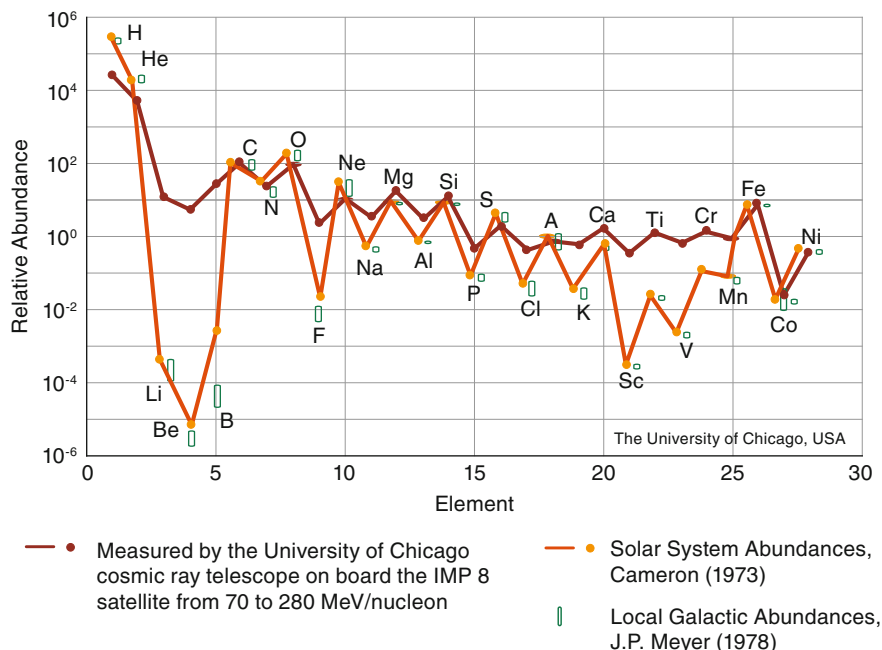


Fig. 5.4-4 The elemental abundance of the cosmic radiation near Earth compared to two compilations of the solar system abundances

To this point, we have discussed the charged particle components of the cosmic radiation. It must be remembered, however, that the supernovae that generated the cosmic rays were prodigious generators of gamma rays and neutrons, as well. Neither is influenced by magnetic fields, and they would therefore travel to Earth along straight-line paths. The known rate of occurrence of supernova (SN) in the galaxy ($10^{-4} \text{ kpc}^{-3} \text{ year}^{-1}$) means that there is a substantial probability (0.32) that a SN has occurred over the past 100,000 years within about 200 pc of Earth. Theoretical considerations (see Sect. 18.3) indicate that a SN at that distance might produce a cosmic gamma ray pulse of sufficient intensity to produce a short-lived (1–10 years), detectable increase in the rate of production of cosmogenic radionuclides on Earth. Although the neutron decays with a half-life of 10 min, the relativistic time dilation (Box 5.3.1) means that a significant number of very high-energy neutrons might survive long enough to initiate highly energetic nucleonic cascades in the atmosphere, also producing cosmogenic radionuclides. The charged cosmic rays from this “neighbourhood SN” would eventually diffuse to the vicinity of the Earth and might yield a long-drawn-out pulse of cosmic rays at Earth if the intergalactic magnetic field from the vicinity of the SN passed near Earth (The rise and fall times of this pulse would be large (~ 10 – 100 times) compared to the time light would take to get here). That is, there is a significant probability that a SN may have occurred close enough to the Earth

in the past 10^5 years for the gamma ray and neutron components to produce a short-lived increase in the production of cosmogenic radionuclides at Earth and a smaller probability that the charged cosmic rays would result in a long-drawn-out pulse in the cosmogenic data. To date, this kind of variability has not been identified in the cosmogenic record; however, it must be regarded as a realistic possibility. Since the neutrons and gamma rays travel in straight lines, the amplitude of any resulting pulse in the cosmogenic record would depend strongly on the declination (i.e., latitude) of the supernova on the celestial sphere. For declinations $>50^\circ\text{N}$ or $>50^\circ\text{S}$, the cosmogenic pulse would be largely restricted to a single hemisphere (see Fig. 13.4-4). One possible candidate for an SN response in the cosmogenic record is discussed in Sect. 18.3.

In summary then, the interstellar intensity of the galactic cosmic radiation near Earth may vary slowly with a time scale of order $\geq 5 \times 10^4$ years. In addition, there is a reasonable probability that a SN may have occurred in our “immediate neighbourhood” (within 200 pc, say) over the past 100,000 years, and that the gamma ray pulse from the SN may have resulted in a significant, short-lived increase in the cosmogenic radionuclide production rate. This is, however, the “exception that proves the rule” that in general, the multiplicity of SN in the galaxy, and the strong trapping imposed by the galactic magnetic field means that the LIS of the cosmic radiation at Earth is relatively constant with time.

5.5 Our Variable Sun

When a cosmic ray approaches to within $\sim 2.25 \times 10^{10}$ km (150 AU) of the Sun, it enters the “heliosphere,” a great bubble of plasma and magnetic fields carved out in the interstellar medium by the Sun. The magnetic fields in the heliosphere have a profound influence on the <20 GeV cosmic rays that generate the majority of the cosmogenic radionuclides and which are detected by neutron monitors and satellite detectors (McDonald 1998). The heliospheric magnetic fields vary substantially from day to day, month to month, year to year, and even century to century, resulting in large fluctuations in the cosmic radiation near Earth. To understand those variations and their impacts on the cosmogenic and instrumental records on Earth, this section provides a brief summary of the manner in which the Sun varies over time. Subsequent sections describe the consequent changes in the heliosphere and the manner in which they “modulate” (i.e., change the intensity) of the cosmic radiation on Earth and throughout the heliosphere.

Over the past several thousand years, mankind has observed the presence of “sunspots” on the face of the Sun (Aschwanden 2005; Eddy 1976). With the advent of telescopes and spectrometers, it became clear that they represent extremely large regions of the Sun’s photosphere in which there are strong, ephemeral magnetic fields (up to 0.3 T in mks units or 3,000 G). Figure 5.5-1 is a photograph of the Sun when several large groups of sunspots were visible; each “sunspot group” may contain >25 individual sunspots. The magnetic fields in a sunspot group change rapidly with time,

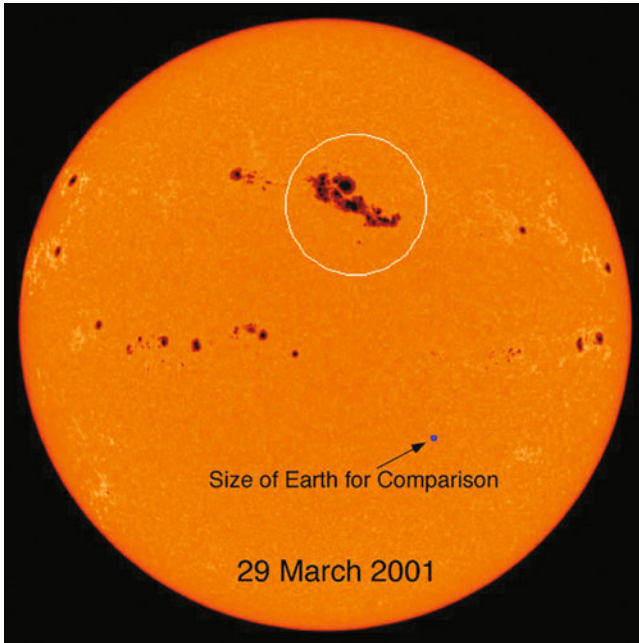


Fig. 5.5-1 The Sun on 29 March 2001 showing several sunspot groups. Each group may contain >25 individual sunspots. The sunspot number (see Box 5.5.1) for this day was 233. The Sun rotates on its axis once every ~27 days as seen from the Earth, a sunspot first appearing on the “eastern limb;” it then crosses the central solar meridian (“central meridian passage,” CMP) 7 days later and then disappears over the west limb (*right hand side*) after another seven days (Martin Ruzek; SpaceWeather.com)

are complex, and extend over a great area (the diameters of the largest sunspot groups are ~30 times the diameter of Earth). Spectrometric measurements show that the magnetic fields associated with sunspots vary greatly from day to day.

5.5.1 BOX The Sun and the Sunspot Number

The *Sun* is a totally undistinguished star in the “main sequence” on the H–R (Hertzsprung–Russell) diagram that plots the dependence of stellar temperature upon luminosity. The rotation of the Sun about its axis is a function of heliographic latitude, being faster at the equator (25 days) and slower at the poles (30 days). The solar material has a high electrical conductivity, and consequently the interactions between the motions of the solar material, and the entrained magnetic fields, are described by the equations of magnetohydrodynamics. In particular, these, and the differential rotation of the Sun lead to the concept that there is a self-reinforcing solar electric dynamo that generates the solar magnetic fields.

The sunspot number is an empirical parameter that is used to describe the degree of solar activity. It has been estimated for each day since 1609, based on the observations made by many individual observers, using the equation $N = k(10g + f)$, where g is the number of sunspot groups and f is the total number of individual spots. The constant k provides inter-calibration between the observatories; this is necessary since the final averaged estimate for the day depends on a varying number of different observatories as a consequence of cloud cover, and particularly in the past, on instrumental factors, and the overall quality of the “seeing” at each particular observatory.

The Chinese and the Koreans were the first to document the presence of sunspots >2,000 years ago. Galileo observed them with a telescope in 1609, and a large number of observers have recorded them ever since. Those >2,000 years of observations show that their number and size vary greatly with time. Figure 5.5-2 shows how an empirical measurement, the “sunspot number,” has varied in the 400 years for which telescopic observations of the Sun have been made (Box 5.5.1). The most obvious feature of these graphs is the strong ~11-year variability, in which the sunspot number increases from near zero to a high value (~100) over a period of ~4–5 years, and then returns to close to zero over the subsequent 6–7 years. This change is quite dramatic; during “sunspot minimum,” there may be no sunspots visible for months at a time, while at “sunspot maximum” only 4 years later, there may be several large sunspot groups on the face of the Sun, containing up to a 100 individual

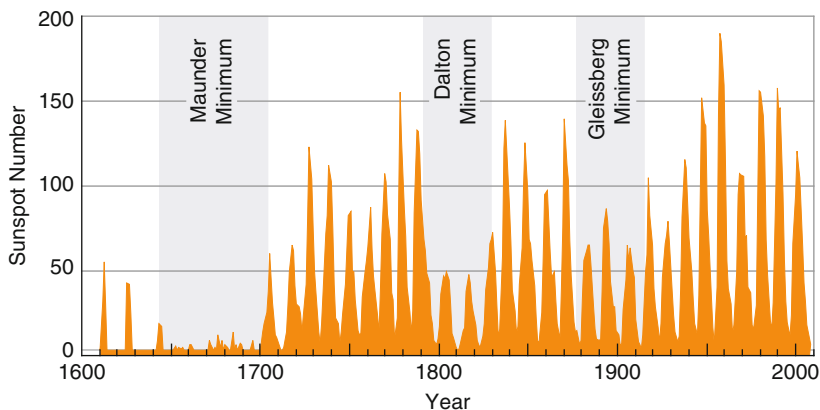


Fig. 5.5-2 The manner in which the monthly average “sunspot number” has varied with time. The solar cycles are numbered: the cycle 1755–1766 is cycle 1, while the last 1996–2007 is cycle 23. Note that the sunspot numbers were low during the intervals marked with “Maunder,” “Dalton,” and “Gleissberg.” These and similar intervals are called the “Grand Minima.” The cosmogenic data indicates that there have been ~26 Grand Minima in the past 9,300 years (Sect. 17.3.2). Present-day estimates suggest that maximum of cycle 24 (2011 AD) will be lower than any in the past 70 years

sunspots, as shown in Fig. 5.5-1. The 11-year solar cycle is sometimes called the “Schwabe cycle,” in honour of its discoverer, Heinrich Schwabe, in 1843 (Schwabe 1844).

Figure 5.5-2 shows that the amplitude of the Schwabe cycle has varied greatly over the past 400 years. For example, starting with the sunspot cycle of 1901–1912, the maximum annual sunspot numbers were 62, 104, 78, 114, 152, and finally 190 for the sunspot cycle 1954–1965. By way of contrast, the sunspot numbers were very low between 1645 and 1715. This long gap is not a consequence of infrequent observations; to the contrary, more than 70 observers were active, frequently resulting in observations for >100 days each year (Hoyt and Schatten 1998). The almost complete absence of sunspots during this interval, called the Maunder Minimum, indicates that the solar magnetic fields were strikingly different from those during the “space age.” The interval 1790–1830, called the Dalton Minimum, also exhibited low sunspot numbers. Extended periods of low solar activity such as these are referred to as the “Grand Minima” in the sunspot record. We will later see that the production rate of the cosmogenic radionuclides shows a strong 11-year variability and long-term secular trends that are associated with the Maunder, Dalton, and other Grand Minima. The physical processes responsible for the 11-year Schwabe cycle and the 22-year Hale cycle, and their secular changes over time, are believed to be due to the operation of a “solar dynamo,” that is not fully understood yet.

Spectrometric methods are used to map the solar magnetic fields, and they show that there is a (roughly) dipole-like “polar” field ($\sim 5 \times 10^{-4}$ T or 5 G) at high solar (“heliographic”) latitudes, in addition to the intense, ephemeral fields associated with the sunspots. Both the polar field and the fields within the sunspots reverse their polarity from one Schwabe cycle to the next. That is, both exhibit a 22-year periodicity called the heliomagnetic or “Hale” cycle. From the point of view of the physics of the Sun, the 22-year periodicity is a dominant feature; however, many of the Sun-induced phenomena at Earth, and in the heliosphere, are largely independent of the polarity of the magnetic fields. As a consequence, the 11-year (Schwabe) periodicity is frequently the dominant feature of geophysical and heliospheric phenomena.

Our forebears, from the Bronze Age onwards, have been fascinated and frightened by the occurrence of solar eclipses. In a total eclipse, when the Moon totally covers the solar disc, “streamers” and “plumes” of light are seen extending up to several solar radii into space above the solar disc (Fig. 5.5-3). These features are due to light scattered from the gas in the vicinity of the Sun; we call this region the “solar corona” (Aschwanden 2005; Eddy 1976). First, drawings and then photographs were made of the corona during the many eclipses that occurred over the past 400 years. Those records provide modern science with an important source of information about the manner in which the corona has varied with time. They show that the nature of the corona has varied strongly from eclipse to eclipse; at a time of low solar activity and low sunspot number, the corona is relatively small and concentrated towards the Sun’s equator but with discernable “dipole-like” plumes of light over the poles of the Sun. When the Sun is active (high sunspot number), the corona is complex; and it extends well away from the solar equator. These observations show that the variations

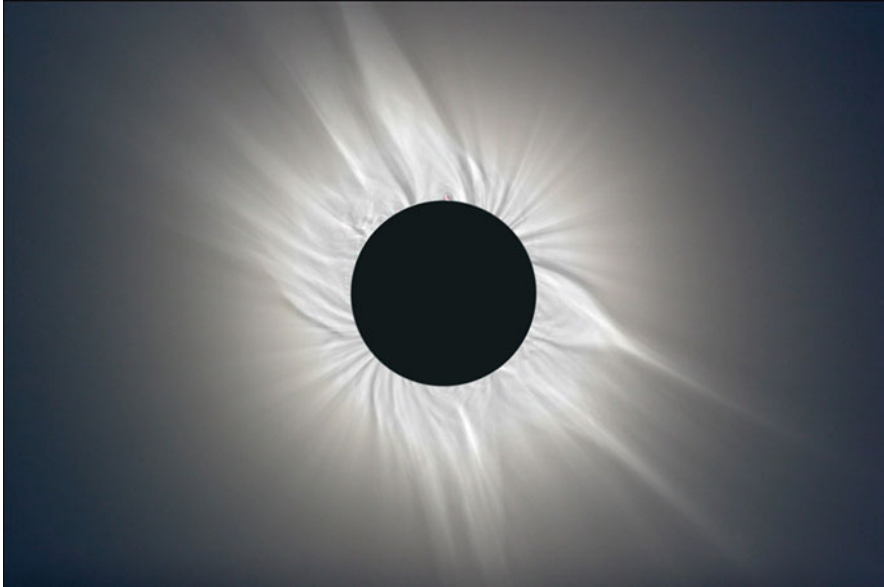


Fig. 5.5-3 The solar corona from a combination of 33 pictures taken during the solar eclipse of March 2006. The faint solar corona is only visible if the bright solar disc is hidden behind the Moon (NASA)

in solar activity have a strong influence on the density of the gas up to 1 and more solar radii above the visible surface of the Sun.

In 1958, Eugene Parker provided a crucial insight into the manner in which the Sun controls the gas density and magnetic field throughout the solar system (Parker 1958, 1963). He showed that the high coronal temperatures imply that coronal gas will stream radially away from the Sun with a velocity of $\sim 300\text{--}500 \text{ km s}^{-1}$. He called this the “solar wind.” The electrical resistance of the high-temperature, low-density solar wind is extremely low, and as a consequence the magnetic energy in the solar plasma cannot be dissipated as kinetic energy (a situation similar to that which causes a superconducting magnet to retain its magnetic field). This means that the solar magnetic field is “frozen in” the solar wind and is then carried with it to the outer reaches of the solar system. As a direct consequence of this and the rotation of the Sun, the solar magnetic field lines are stretched out into the interplanetary medium in the form of an “Archimedes spiral,” as shown in Fig. 5.5-4. This field configuration is sometimes called the “Parker spiral field.” This will be discussed in more detail in the next section.

The space age has allowed us to make continuous observations of many of the properties of the corona and the solar wind. In particular, the flight of the “Skylab” manned space laboratory in 1973 provided an important insight into the nature of the corona itself, the “coronal plumes” seen during eclipses, and the 27-day recurrence of geomagnetic storms. It revealed the presence of lower temperature regions of the corona, called “coronal holes,” that are now known to be associated with a number of important features of the heliosphere. They are the source of a fast

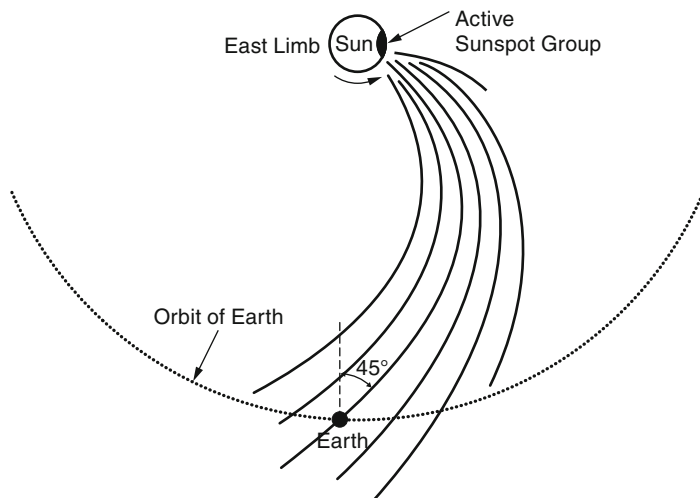


Fig. 5.5-4 The Parker spiral magnetic field. Kinks and other small-scale irregularities in the magnetic field are superimposed on this smooth configuration as discussed in this and the next section. The number and intensity of these irregularities is lowest at sunspot minimum, increasing in broad correlation with solar activity

solar wind ($\sim 800 \text{ km s}^{-1}$), while the solar wind from the regions outside the coronal holes has a speed of $300\text{--}500 \text{ km s}^{-1}$. Further, the magnetic field lines from the chromosphere in the coronal holes connect directly to the interplanetary field and extend out beyond the orbit of Earth to the Termination Shock (Sect. 5.6); this is called the “open” magnetic field of the Sun. The field lines are always “open” above the solar poles. In reality, the lines of force of the “open” field must ultimately return to the Sun; however, they do so from the outer heliosphere.

In the closed field regions (such as above a sunspot group), the magnetic field lines are in the form of loops with both ends anchored in the chromosphere. These regions are the source of the “slow” solar wind ($\sim 300\text{--}500 \text{ km s}^{-1}$). During contemporary sunspot minima, there are large coronal holes in the polar regions of the Sun, extending in parts towards the equator. In the years following sunspot minimum, more and more magnetic loops form in association with the sunspots resulting in large regions of closed field, while the regions of open field shrink until they are largely confined to the poles at sunspot maximum.

X-ray detectors and imaging coronagraphs on satellites have revealed that the corona is a very dynamic part of the Sun, and that the ionized, electrically neutral gas is strongly controlled by magnetic forces. The observations show that a low level of disturbance continues during sunspot minimum. Then, as the number of sunspots increases, a number of violent phenomena begin to occur (the Sun is said to become active), and these make important contributions to the properties of the solar wind and the interplanetary magnetic field. For example, an instability may form in a closed magnetic loop above an active region, leading to reconnection of oppositely directed lines of force and the instantaneous release of very large

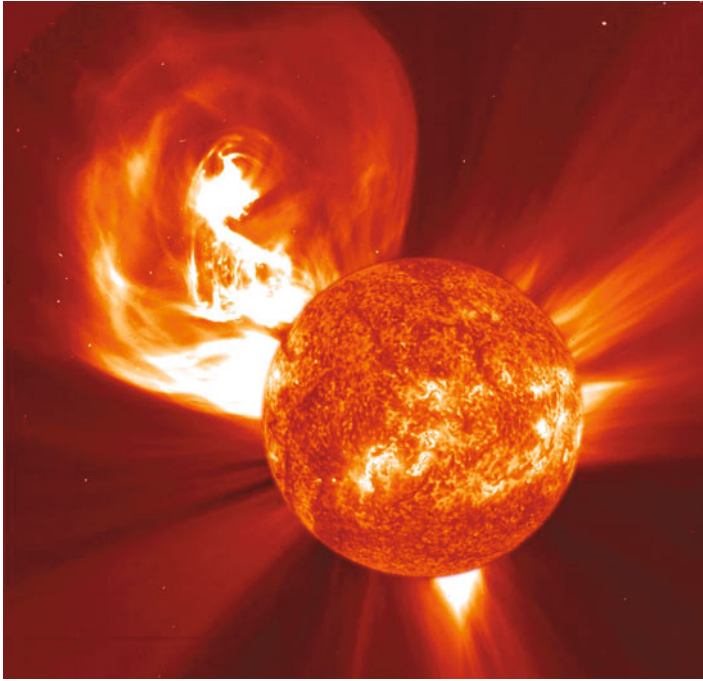


Fig. 5.5-5 The coronal mass ejection observed on January 4, 2002 (NASA)

amounts of magnetic energy. This heats the gas, resulting in a strong eruption that can blow away millions of tons of plasma at speeds of $500\text{--}2,500\text{ km s}^{-1}$; these are called “coronal mass ejections” (CMEs) and an example is given in Fig. 5.5-5.

A fast CME will initiate a powerful supersonic shock wave in the interplanetary plasma that travels outwards from the Sun, sometimes resulting in magnetic storms at Earth (Sect. 5.8.3), and also reducing the intensity of the cosmic radiation at Earth (Sect. 7.4). As in the case of the supernova discussed in Sect. 5.4, these shock waves can accelerate cosmic rays (but to a much lower energy than in the case of the supernova). We will discuss these “solar cosmic rays” and their impact on Earth in Chap. 8.

“Solar flares” are another spectacular form of solar event (Fig. 5.5-6). They appear as a sudden (a rise time of $<1\text{ min}$), intense brightening of the light emission in the vicinity of sunspots, and a great deal of magnetic energy is released within minutes. Large volumes of the chromosphere and corona are heated to high temperatures: copious fluxes of ultraviolet, gamma-, and X-rays, and radio waves are emitted, and on rare occasions, high-energy ($>1\text{ GeV}$) cosmic rays. They are frequently associated with a CME as described in the previous paragraph.

At sunspot minimum, the solar wind flows away from the Sun in a relatively calm, ordered manner. There are few CMEs or other changes in the solar wind

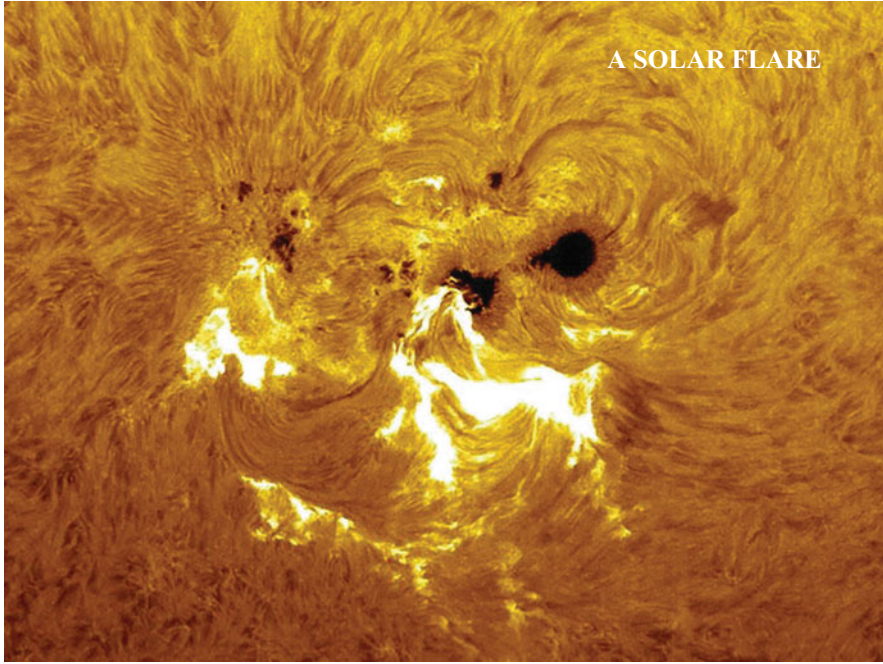


Fig. 5.5-6 A solar flare in the vicinity of a large sunspot group. The annihilation of a portion of the sunspot magnetic fields high in the corona has released a great deal of energy, part of which has heated the gas lower down in the chromosphere and photosphere, leading to the immediate emission of visible and ultraviolet light (NASA)

speed. As a consequence, the interplanetary magnetic field is relatively smooth with only a few kinks or other irregularities in the lines of force. With increasing solar activity, the frequency of occurrence of CMEs increases rapidly (by a factor of 10–20 by the time sunspot maximum is reached). The shock waves they produce, and the interaction regions between the fast wind from coronal holes and the rest of the solar wind, generate turbulence in the solar wind. Because the magnetic field is “frozen-in” to the wind, this turbulence results in the generation of kinks and irregularities in the interplanetary magnetic field. Throughout the solar cycle the interplanetary field retains the general large-scale characteristics of the Parker spiral, but with a steadily increasing number of irregularities up to solar maximum, decreasing thereafter. The strength of the interplanetary magnetic field also varies throughout the solar cycle, being weakest at sunspot minimum, and strongest near sunspot maximum. We will see in Sect. 5.7 that these changes in the irregularities in the interplanetary field and its field strength reduce the intensity of cosmic rays at Earth and result in a large reduction in the production rates of the cosmogenic radionuclides.

Finally, we sound a note of caution. The space age (1957–present) has provided an enormous increase in our knowledge of the Sun. It is easy to fall into the trap of

assuming that this body of knowledge has applied at all earlier times. Note however that Fig. 5.5-2 shows that the peak sunspot numbers for the majority of the space age have been the greatest in the past 400 years. It is important to recognize that the solar phenomena, and the properties of the heliospheric magnetic field may, we stress, may, have been significantly different during the Grand Minima, such as the Maunder Minimum, and the periods in between. The cosmogenic radionuclides provide us with the means to investigate this possibility (Chap. 17).

5.6 The Heliosphere, the Termination Shock, and the Current Sheet

To an observer on Earth, the surface of the Sun seems to be very well defined. Indeed, the visible light is emitted from a very distinct layer of only a few hundred kilometres thickness, called the photosphere, which occupies less than a thousandth of the solar radius. In reality, there is no sharp boundary of the Sun. The density decreases exponentially with increasing height above the photosphere and the temperature changes only slightly throughout this region called the chromosphere. At a height of 2,000 km above the photosphere the density decreases by a factor of ~ 100 and the temperature increases correspondingly from about 10,000 K to 1,000,000 K. This abrupt change marks the transition from the chromosphere to the corona discussed in the previous section. The reason for this dramatic change in temperature is not understood in detail but is believed to be due to the transport of magnetic energy from the region below the photosphere by “magnetohydrodynamic waves,” a form of wave motion that occurs in material with a very low electrical resistance.

As discussed in Sect. 5.5, satellite and other observations have confirmed Eugene Parker’s prediction that there would be a “solar wind” streaming away from the corona. Throughout the period of high solar activity that has coincided with the space age, velocities have been observed to lie in the range $300\text{--}900\text{ km s}^{-1}$. The Sun rotates about its axis once every 25 days, while the solar wind leaves the Sun in a radial direction. As discussed in Sect. 5.5, the plasma streaming from a given point on the Sun forms a spiral pattern (known by mathematicians as an “Archimedes spiral”). Since the solar magnetic field from the given point is entrained in the plasma, the field lines are Archimedes spirals also. Typically, near the orbit of Earth, they make an angle of about 45° with the line from the point of observation to the Sun (Fig. 5.5-4).

The solar wind, with its entrained magnetic field, expands outwards for about a year until it reaches the point at which its pressure equals that of the local interstellar medium (gas plus magnetic field). At this point, a strong standing shock wave develops, called the “termination shock,” as shown in Fig. 5.6-1. Outside the termination shock, there is a thick region of more slowly moving solar plasma, called the “heliosheath,” and beyond that, the “heliopause” and possibly the “bow shock,” which define the interface between solar and interstellar

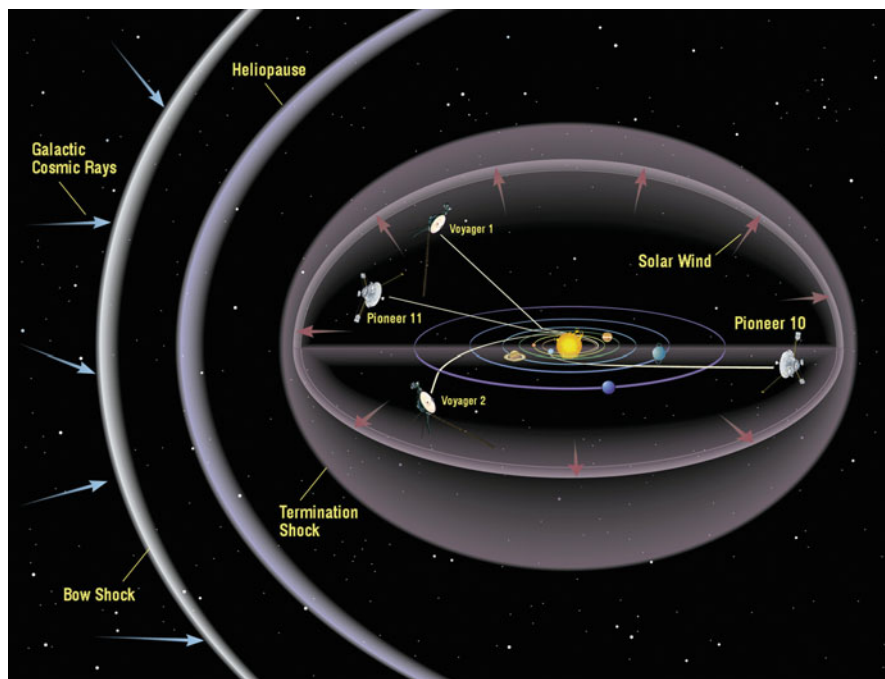


Fig. 5.6-1 A schematic diagram of the major features of the heliosphere. The magnetic field is in the form of the “Parker spiral field” until the “Termination Shock” is reached at $\sim 90\text{--}95$ Astronomical Units (AU). Beyond that is the “Heliosheath” until the “Heliopause” is reached and beyond that (possibly) the “bow shock” at ~ 150 AU that separates the heliosphere from the interstellar medium. The asymmetric shape of the heliosphere is the result of the Sun’s motion (to the *left*) with respect to the interstellar medium. The figure also shows the orbits of Pioneers 10 and 11, and Voyagers 1 and 2, that have provided a great deal of information about the three-dimensional nature of the heliosphere (NASA)

material. The caption of Fig. 5.6-1 uses the astronomical unit (AU) as the unit of length (see Sect. 5.3). The AU is commonly used in studies of the solar system, and the heliosphere, and is defined as the average distance of the Earth from the Sun ($1 \text{ AU} \cong 1.5 \times 10^8 \text{ km}$).

Satellite measurements of the density and speed of the solar wind made since 1962 predicted that the termination shock would be in the vicinity of $85\text{--}95$ AU from the Sun. This was verified in 2004 when Voyager 1 passed through it at 94 AU and then in 2007 when Voyager 2 crossed at a higher latitude and 84 AU from the Sun. Inside the termination shock, the heliospheric magnetic field is relatively well ordered in the shape of Archimedes spirals, upon which are superposed magnetic irregularities, as a consequence of turbulence in the solar wind (see Sect. 5.5). In Sect. 5.7, we will see that these magnetic fields have a strong effect upon the cosmic rays that reach the Earth, leading to substantial variations in the counting rates of cosmic ray instrumentation and in the rate of production of the cosmogenic radionuclides.

It is possible that the termination shock was closer to, or further away from, the Earth in the past. As noted above, its distance from the Sun is determined by the pressure of the local interstellar material (i.e., outside the bow shock), and astronomical studies suggest that this has varied in the past. Thus, the Earth (and heliosphere) is now in a relatively warm, medium-density interstellar gas cloud, having entered it from a colder, less dense interstellar cloud sometime in the past 100,000 years. The lower pressure of the interstellar gas at that time would mean that the termination shock was further away from the Sun than in the present epoch. Alternatively, a substantial reduction (say) in the magnetic field entrained in the solar wind, or the density of the solar wind would reduce its pressure on the interstellar medium, and the termination shock would be closer to the Sun. The possible consequences of this variability upon the production of cosmogenic radionuclides are discussed in Chap. 18.

The earliest satellite measurements of the heliospheric field by Norman Ness and his colleagues in 1963 revealed a surprising result: the polarity of the magnetic field near Earth would switch from outward pointing, to inward pointing and back every 27 days (i.e., the solar rotation period). More surprising still, the switch occurred very quickly, frequently <1 h. This was initially called the interplanetary sector structure. Those and all subsequent observations of a similar nature are now interpreted in terms of the “heliospheric current sheet” (also known as the “neutral sheet”): a thin surface carrying an electrical current that separates the outward pointing fields in one hemisphere from the inward pointing fields in the other. Figure 5.6-2 is an artist’s representation of the current sheet. It is remarkably thin, about 10,000 km, which is comparable to the gyroradii of the <10 GeV cosmic rays



Fig. 5.6-2 An artist’s view of the heliospheric current sheet. We are looking down on the solar system at an angle of about 45° with the Sun at the centre. The magnetic fields above the current sheet are “Parker spirals” (Fig. 5.5-4) of one polarity; those below are also Parker spirals of the opposite polarity. Both polarities reverse near the maximum of each solar cycle (Wikipedia)

that produce the majority of the cosmogenic radionuclides of interest to us. As we will see in the next section, this means that the current sheet provides a “channel” that opens a relatively easy passage for the cosmic rays in some directions through the heliosphere, and it therefore has an important influence on the modulation of the galactic cosmic radiation.

5.7 Modulation of the Cosmic Radiation in the Heliosphere

The previous sections have summarized how the magnetic properties of the Sun vary with time, leading to short- and long-term variations in the strength and irregularities in the heliospheric magnetic field. In this section, we discuss the manner in which these changes in the magnetic field “modulate” (or vary) the cosmic ray intensity near Earth.

Theoretical and experimental studies of the cosmic ray modulation have been in progress for about 75 years, and it is still an active field of research. This presents us, the authors, and you, the reader, with a problem. Namely, there are a number of different formulations of the modulation process in the literature that cause much confusion even within the cosmic ray research community. In broad terms, this multiplicity breaks down into two different groups: (1) descriptions based on the accurate but difficult-to-use “cosmic ray propagation equation” (Sect. 5.7.1); and (2) descriptions based on a quasi-empirical approximation called the “modulation function” (Sect. 5.7.2). Further differences are then introduced to both groups by the existence of a number of different formulations of the local interstellar cosmic ray spectrum. As a consequence, you, the reader, will find somewhat different values in the literature and on the Internet for the modulation function for the same moment in time. This can be unsettling to even the most skilled cosmic ray research scientist.

Our approach has been, first, to provide general outlines of the propagation equation and modulation function approaches. Our aim here is to provide some familiarity with the terminology and the physics underlying both methodologies, without going deeply into the mathematics. Then, in the case of the modulation function approach, we have selected a commonly used formulation for our discussions throughout this book. Further, as we discuss in Sect. 5.7.4, we have used one estimate of the LIS for all our discussions and in the preparation of all our figures. We have then provided a conversion graph that allows conversion to and from modulation functions based on other LIS. It is stressed that the modulation function is only an approximation, and the lack of exact agreement between the various formulations is the inescapable consequence of the different data and assumptions they use. Each formulation is essentially as good as another; however, any investigation of the cosmogenic data should be in terms of one formulation (and one assumed LIS) alone.

5.7.1 The Cosmic Ray Propagation Equation

On entering the heliosphere (Fig. 5.6-1), a cosmic ray experiences a very different world from the one in interstellar space. In the heliosheath, the environment is not so different – the magnetic fields are somewhat stronger and more disordered than in interstellar space. The solar plasma is expanding away from the Sun at the rather languid speed (for solar physics) of $\sim 100 \text{ km s}^{-1}$ and carrying the magnetic field with it. These fields have a substantial influence on low-energy ($< 100 \text{ MeV/nucleon}$) cosmic rays but (as far as we can tell at present) a negligible effect on the cosmic rays that generate the cosmogenic radionuclides on Earth, or are detected by neutron monitors (Caballero-Lopez et al. 2004; Jokipii 1991; Parker 1963).

The real difference is seen when the cosmic ray passes through the termination shock and starts travelling towards the Sun. The magnetic field steadily gets stronger; by the time it reaches the orbit of Earth, it has increased by a factor of ~ 500 . As a consequence, the gyroradii of the cosmic rays decrease by the same factor, and they travel in an increasingly tight spiral motion. Turbulence and shock waves in the solar wind caused by CME, and other phenomena on the Sun, result in small-scale distortions of the magnetic field as discussed in Sect. 5.5. These distortions deflect the cosmic ray in a random manner that can be described mathematically as a diffusion process. The distortions in the magnetic field are being swept outwards from the Sun at the solar wind speed ($300\text{--}900 \text{ km s}^{-1}$), and in being deflected by the magnetic irregularities, an incoming cosmic ray is decelerated by a small amount. The spiral nature of the underlying magnetic field results in “drift effects,” which cause the gyrating cosmic ray to move in a direction at right angles to the direction of the Parker spiral field (see Sect. 5.7.5). The drift direction reverses when the polarity of the magnetic field changes. The overall consequence of these several processes is that the cosmic ray intensity declines steadily in going from the termination shock, towards the Sun. The process whereby the solar wind and the entrained magnetic field reduce the cosmic ray intensity inside the heliosphere is referred to as “cosmic ray modulation.”

The combined effect of these processes is described by the “cosmic ray transport equation,” which is a specialized application of the Fokker–Planck equation (Parker 1963).

$$\frac{\partial U}{\partial t} = \nabla \kappa \nabla U - \nabla(VU) - v_D \nabla U + \frac{1}{3} \nabla \cdot \nabla \frac{\partial}{\partial T} (\alpha T U) \quad (5.7.1-1)$$

Diffusion Convection Drift Deceleration

where $U(x, E, t)$ is the cosmic ray number density per unit kinetic energy interval E at time t , κ is the diffusion tensor, V is the solar wind speed, v_D is the effective drift speed in the mean magnetic field (discussed below), and $\alpha(E) = (E + 2E_0)/(E + E_0)$, where E_0 is the particle rest mass.

One-dimensional diffusion is often discussed using the concept of “mean free path,” λ , instead of the diffusion coefficient, κ , where $\kappa = (1/3) \lambda v$ (where v is the particle velocity). This has the advantage that λ can be understood in physical terms as the average distance the particle travels before it is scattered. Thus, it is easy to visualize that the mean free path, λ , decreases as the number of scattering centres increases, and it becomes more difficult for the particle to make progress. From the relation $\kappa = (1/3) \lambda v$, it is clear that a small diffusion coefficient corresponds to a small mean free path, which implies a strongly scattering medium, while a large κ corresponds to little diffusion. We will use this relationship between mean free path and diffusion coefficient to assist the understanding of the following discussion.

We now outline the physical processes described by the four terms in the cosmic ray transport equation. Experience has shown us that many experimentalists without a background in mathematics, or cosmic ray physics, find it hard to understand the interaction and function of the several mathematical terms in the equation. Consequently, to assist in a conceptual understanding of the overall modulation process, we will use the analogy of a lot of fish swimming up a fast-flowing, rocky, stream. Cosmic ray theoreticians will hate the analogy, but we believe it will be helpful to many others.

The First Term. In the cosmic ray transport equation, the first term describes the diffusion of the cosmic ray particles through the irregularities in the magnetic field that are “frozen-in” the outward moving solar wind. The diffusion tensor k specifies the scattering properties of the heliospheric magnetic field, and since the gyroradius of a cosmic ray (Eq. 5.3-6a) increases as a function of rigidity (and therefore momentum and energy) and decreases as the reciprocal of the magnetic field strength, the tensor terms depend on both the direction and strength of the magnetic field and also on the energy of the cosmic ray. Satellite measurements have shown that the monthly average interplanetary magnetic field strength varies by a factor of 1.5–2.0 throughout the solar cycle (from ~5 to ~10 nT near Earth since 1960) and that the number of CMEs (and hence the turbulence and the number of magnetic irregularities) changes by a factor of 10–20. As a consequence, the mean free path, and hence the diffusion coefficients, are large during sunspot minimum, then become significantly smaller with the onset of solar activity following sunspot minimum, with the result that the cosmic rays find it much harder to approach the Sun. That is, this term describes the manner in which the solar wind, plus its entrained magnetic field, makes it difficult for the cosmic rays to simply spiral along the magnetic lines of force into the inner solar system.

In the analogy, as they swim up the river, the fish keep swimming into the rocks, and water eddies, which divert them from where they want to go upstream. Some may even go in the wrong direction for a while – across the river – even downstream. They only manage to keep going up the stream quite slowly against the current and the rocks. That is, they “diffuse” up the stream, rather than going purposefully upwards.

The Second Term. The second term in (5.7.1-1) represents the manner in which the solar wind “convects” the cosmic rays out of the solar system. Thus, each

cosmic ray, when scattered by an irregularity in the magnetic field, has imparted to it an outward velocity. As a consequence, on average, all the cosmic rays entering the heliosphere are being pushed (“convected”) outwards away from the Sun. Examination of this term in the equation shows that it depends linearly upon the speed of the solar wind, and we can therefore anticipate that the cosmic ray modulation will be greater (less cosmic rays near Earth) when the solar wind speed is high.

In the analogy, some of the fish never recover from being deflected by the rocks and eddies, and they are swept away downstream, and never come back. Even if going straight upstream, a fish is fighting against the flow, and it is slowed down. The faster the stream, the more fish are swept downstream, and the number of fish gaining their objective is less (lower intensity of fish).

The Third Term. The third term of the cosmic ray transport equation specifies the manner in which both the gradients and curvature of the underlying heliospheric magnetic field give rise to second order effects, called the “drift effects” (Jokipii et al. 1977). The drift velocity vector, v_D is a function of the magnetic field strength, its gradient, and the curvature of its field lines. (The drift effects will be discussed in more detail in Sect. 5.7.5). For the purpose of this discussion, we merely note that the drift directions and the direction of the vector v_D in (5.7.1-1) (and therefore the contribution to the modulation) depend upon the polarity of the heliospheric magnetic field. The 22-year “Hale cycle” in the polarity of the solar magnetic fields therefore introduces a 22-year periodicity into the “11-year” modulation observed by neutron monitors and satellite detectors. Presently, the consequences of these drift effects are barely visible in the cosmogenic record; however, this will change as the statistical errors and system effects are reduced in the near future.

In the analogy, it so happens that the bears that want to eat the fish are very intelligent bears. They have agreed to move from one bank of the river to the other every solar cycle. This forces the fish across the river in one direction during one solar cycle, in the other direction during the following solar cycle. There are more rocks on one side of the river than the other, so the “modulating effects” on the fish vary from one solar cycle to the next.

The Fourth Term. The fourth term in (5.7.1-1) describes the adiabatic deceleration of cosmic ray particles, in which the energy of the individual cosmic rays decreases as they spiral inwards and see a steadily increasing magnetic field and when they scatter on the magnetic inhomogeneities moving with the solar wind. Consequently, this process depends both on the cosmic ray energy and the speed of the solar wind.

In the analogy, the fish lose a lot of energy as they fight against the rocks, eddies, and current, and they are very tired when some finally reach their destination.

The total effect then is that the number of cosmic rays reaching the inner solar system is reduced by convection and in addition the energy of the particles is reduced. Because of the latter effect, an observer looking at galactic cosmic rays of energy E near Earth sees particles that were of higher energy, $E + \Delta E$, in interstellar space. The cosmic ray spectrum decreases steeply towards higher energies (Fig. 5.4-3) and consequently the intensity outside the heliosphere

decreases as ΔE increases. That is, as the deceleration process increases during the start of the solar cycle, the portion of the LIS sampled by the observer shifts to higher energies and the cosmic ray intensity at Earth decreases.

The cosmic ray transport equation cannot be solved analytically and is difficult to use in the form given above. Using detailed mathematical models based on the transport equation; however, Jokipii and his co-workers (Jokipii 1991; Jokipii et al. 1977) and the South African school (Potgieter and Moraal 1985; Potgieter and Leroux 1992) have shown that the transport equation accounts for the observed features of the cosmic ray intensity inside the heliosphere. It is the yardstick against which approximations such as the modulation function (see below) are to be judged.

For our purposes in using cosmogenic and other cosmic ray data, it is usually sufficient to use the “modulation function” or an allied quantity, the “modulation potential” as an approximation to the cosmic ray transport equation. They include all of the processes other than the drift effects (term 3) and provide a useful relative description of the cosmic ray modulation for energies >500 MeV. They are described in Sect. 5.7.3. They are already used in practical applications, which require a parametric measure of the intensity of the galactic cosmic radiation. Such applications include the prediction of radiation damage to spacecraft and the prediction of the radiation dose that astronauts and passengers in high-flying aircraft will receive. Since they provides a single parameter that quantifies the cosmic ray and solar activity in the past, the authors expect that they will soon be used in archaeology, climate studies, and other disciplines that require a quantitative, easy-to-use measure of the solar influence on the cosmic radiation intensity at Earth in the past.

In summary then, the intensity of the 0.5–20 GeV cosmic radiation is reduced as a consequence of their propagation through the magnetic field embedded in the outward-flowing solar wind. The presence of irregularities in the magnetic field is the major contributor to this process, and as the number of irregularities, and the strength of the field increases rapidly towards sunspot maximum, the cosmic ray intensity decreases at Earth. As the solar activity declines after sunspot maximum, the reverse applies, and the cosmic ray intensity increases to a maximum value at sunspot minimum.

5.7.2 *The Local Interstellar Spectrum*

Section 5.4 has introduced the concept of the “local interstellar spectrum”—the cosmic ray spectrum that would be observed outside the heliosphere. Clearly, the spectrum observed near Earth will be determined, in large part, by the intensity and energy (or rigidity) dependence of the LIS. Unfortunately, we have no direct measurements of the LIS yet, and must depend on estimates based on measurements of the spectrum near Earth.

It is beyond the scope of this book to discuss the details of the estimation process [e.g., (Webber and Lockwood 2001b)]. Suffice to say, it starts with the theoretical prediction that the supernova acceleration process (Sect. 5.3) yields an E^{-2} differential energy spectrum. Spallation interactions (see Box 10.3.1.1) with interstellar matter and leakage of the higher energy (see Eq. 5.3-6a) particles from the galaxy result in a steeper LIS. It is usually assumed then that there is negligible solar modulation above 20 GeV, and the spectrum above that energy observed by satellites and high-flying balloons is assumed to represent the LIS. Various methods have been used to estimate the LIS below 20 GeV, always based on extrapolation, with various forms of control based on the measured spectrum near Earth.

In addition to differences due to different extrapolation methods, the accuracy of the spectra observed near Earth has increased over the past decades, and observations have been made in the outer heliosphere where there is less solar modulation. Further, analytical approximations have been developed to provide simplicity in mathematical models. As a consequence of all three factors, a number of different formulations of the LIS will be found in the literature. More will be formulated in the future. Clearly, each formulation of the LIS implies a slightly different amount of modulation to yield the observed GCR intensities near Earth (or the observed production rates of the cosmogenic radionuclides).

To date, little has been done to reduce the confusion created by this degree of complexity and the several approaches to the modulation function that are outlined in the next section. Briefly, the LIS formulations that are most commonly encountered in the cosmogenic literature are:

1. Castagnoli and Lal (1980). This was based on the early satellite data of Garcia-Munoz et al. (1975) and has been commonly used in both the glaciological and cosmogenic literature since 1980.

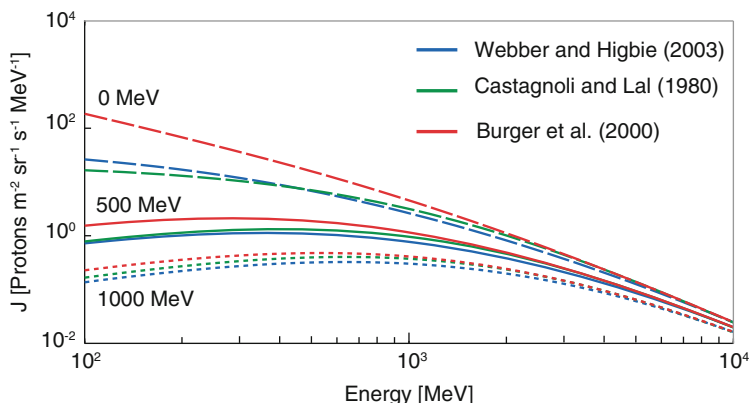


Fig. 5.7.2-1 Three estimates of the Local Interstellar Spectrum (LIS) of the proton component of the cosmic radiation ($\Phi = 0$ MeV) and those spectra after modulation quantified by modulation functions (a) $\Phi = 500$ MeV similar to sunspot minima conditions since 1954; and (b) $\Phi = 1,000$ MeV, similar to sunspot maximum conditions since 1954. After Steinhilber et al. (2008)

2. Webber and Lockwood (2001b) and Webber and Higbie (2003). These use recent satellite measurements of the cosmic ray intensity in the outer heliosphere, thereby minimizing the effects of cosmic ray modulation.
3. Others as used in recent years by Usoskin, Scherer, and the South African School [e.g., Burger et al. (2000)].

Box 5.7.2.1 provides more detail on these and other estimates of the LIS spectra. Figure 5.7.2-1 compares the proton spectra for three of these. Note that the first two are in reasonable agreement and that the greatest differences between all three occur for $E < 1$ GeV.

Nevertheless, the multiplicity of LIS spectra introduces uncertainties into the study of the cosmic ray modulation and into the estimation of the “modulation function” discussed in the next section. To reduce this confusion, Sect. 5.7.4 states the convention with regard to the LIS used throughout this book, and Fig. 5.7.3-3 provides the means to compare the values of the modulation function obtained using different LIS formulations.

5.7.2.1 BOX Estimates of the Local Interstellar Spectrum (LIS)

Herbst et al. (2010) compiled parameterizations of estimates of the differential LIS referred to most frequently in the cosmogenic and allied literature, which are as follows. The units for all LIS are $(\text{particles m}^{-2} \text{ sr}^{-1} \text{ s}^{-1} (\text{MeV/nucleon})^{-1})$.

Garcia-Munoz et al. (1975):

$$J_{\text{LIS}}(E) = A[E + B \exp(CE)]^{-\gamma}$$

where E is MeV/nucleon and

for protons, $A = 9.9 \times 10^8$, $B = 780$; $C = -2.5 \times 10^{-4}$, and $\gamma = 2.65$.

for helium, $A = 1.8 \times 10^8$, $B = 660$; $C = -1.4 \times 10^{-4}$, and $\gamma = 2.77$.

Castagnoli and Lal (1980):

In the first comprehensive study of the production of the cosmogenic radionuclides by the modulated cosmic rays, Castagnoli and Lal assumed that the helium and heavier cosmic ray spectra (Garcia-Munoz et al. 1975) have the same spectral shape as the proton spectrum.

Masarik and Beer (1999) used the Castagnoli and Lal spectrum for the protons, helium, and heavies, while publishing (in error) the exponent in the power law as $\gamma = 2.5$. This error was noted in Masarik and Beer (2009) and we confirm that all the calculations in the earlier work by Masarik and Beer (1999) used $\gamma = 2.65$. That is, all the computations in this book and in all the literature based on the Masarik and Beer computations are based on the proton spectrum quoted by Garcia-Munoz et al. (1975), but with allowance being made for the helium and heavy components of the LIS. The continued use of this old spectrum is discussed in Sect. 5.7.4.

Usoskin et al. (2005) have approximated the Burger et al. proton spectrum (Burger et al. 2000):

$$J_{\text{p-LIS}}(E) = \frac{a}{b[E(E + 2E_0)]^{1.39} + c[E(E + 2E_0)]^{0.135}}$$

where $a = 415.7$, $b = 10^{-7}$, and $c = 1.6488$.

Webber and Higbie (2003):

This reference restated the spectrum of Webber and Lockwood (2001a) as follows:

For protons $J_{\text{p, LIS}}(E) = (21.1 E^{-2.8}) / (1 + 5.85 E^{-1.22} + 1.18 E^{-2.54})$

For helium $J_{\text{He, LIS}}(E) = (1.075 E^{-2.8}) / (1 + 3.91 E^{-1.22} + 0.90 E^{-2.54})$

where E is MeV/nucleon and the flux units are particles/(m² sr s MeV/nucleon).

Langner et al. (2003):

This LIS is a parametrization based on the GCR propagation model of Moskalenko et al. (2002).

$$J_{\text{p-LIS}}(E) = \exp\left(a - b[\ln E]^2 + c \ln E - d\sqrt{E}\right) \text{ for } E < 1,000 \text{ MeV/nucleon}$$

$$J_{\text{p-LIS}}(E) = \exp\left(e - f \ln E - \frac{g}{E}\right) \text{ for } E \geq 1,000 \text{ MeV/nucleon}$$

with $a = 0.823$, $b = 0.08$, $c = 1.105$, $d = 9.202 \cdot 10^{-2}$, $e = 22.976$, $f = 2.86$, and $g = 1,500$.

Webber and Higbie (2009):

$$J_{\text{p-LIS}}(E) = \exp\left(a + b \ln[\ln E]^2 + c\sqrt{\ln E} + \frac{d}{\ln E} + \frac{e}{(\ln E)^2}\right)$$

The parameters in this expression depend on the energy as follows:

Parameter	$E < 1,000 \text{ MeV/nucleon}$	$E \geq 1,000 \text{ MeV/nucleon}$
a	-124.4	0
b	-51.83	-51.68
c	131.6	103.5
d	-241.7	-709.7
e	376.6	1,161

5.7.3 The Cosmic Ray Modulation Function and Potential

The cosmic ray transport equation is quite general and can be used (with difficulty) to calculate the three-dimensional behaviour of the cosmic radiation throughout any assumed model of the heliosphere. For our purposes, however, we are only

interested in the cosmic radiation that reaches Earth or the inner solar system <10 AU (for meteoritic studies). An important step towards providing a more practical quantification of the modulation in this case was provided by Leo Gleeson and Ian Axford in 1968 when they developed the concept of the “modulation function” (Caballero-Lopez and Moraal 2004; Gleeson and Axford 1968).

It is important to note that the modulation function (and modulation potential; see later) is based on a number of simplifying assumptions. They are (1) that the heliosphere is spherically symmetric; (2) that it is only changing slowly (order of a month or more); (3) that the dependence of the cosmic ray flux on direction (“the anisotropy”) is small; and (4) that the diffusion effects can be expressed as separable functions of rigidity and distance from the Sun. The first three assumptions are valid for annual average cosmogenic data and monthly or longer averages of neutron monitor or satellite data. Those simplifications, and several others that are borne out in practice, allow the transport equation to be simplified thus:

$$J_T(T, \Phi) = J_{\text{LIS}}(T + \Phi) \frac{(T^2 - E_0^2)}{(T + \Phi)^2 - E_0^2} \quad (5.7.3-1)$$

where

$J_T(T, \Phi)$ is the differential cosmic ray flux at Earth in particles $\text{m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{MeV}^{-1}$, corresponding to a specified value of the modulation function, Φ . Here,

$J_{\text{LIS}}(T)$ is the differential local interstellar spectrum (cosmic ray flux outside the heliosphere) in the same units.

T is the total energy of the particle (rest mass plus kinetic energy) in MeV. For alphas and heavies, this is AT_p where A is the atomic weight of the cosmic ray, and T_p is the total energy of a proton.

E_0 is the rest mass energy in MeV for all nucleons as in the above comment. For a single nucleon $E_0 = 938 \text{ MeV}$

and Φ is called the “modulation function” and is measured in MeV.

Frequently, Eq. (5.7.3-1) is rewritten in terms of kinetic energy, E , and rest mass energy E_0

$$J(E, \Phi) = J_{\text{LIS}}(E + \Phi) \frac{E(E + 2E_0)}{(E + \Phi)(E + \Phi + 2E_0)} \quad (5.7.3-1a)$$

Equations (5.7.3-1) and (5.7.3-1a) apply to the proton, alpha, and heavier components of the cosmic radiation; however, the modulation function Φ for a given point in time varies from one component to another, as discussed below.

Inspection of (5.7.3-1a) shows that the cosmic ray intensity at Earth, for a given kinetic energy, E , is given by the product of the LIS intensity at an energy Φ MeV higher, and another function of kinetic energy, E , E_0 , and Φ . It can be understood in general terms as if the cosmic ray LIS spectrum at energy $(E + \Phi)$ had been

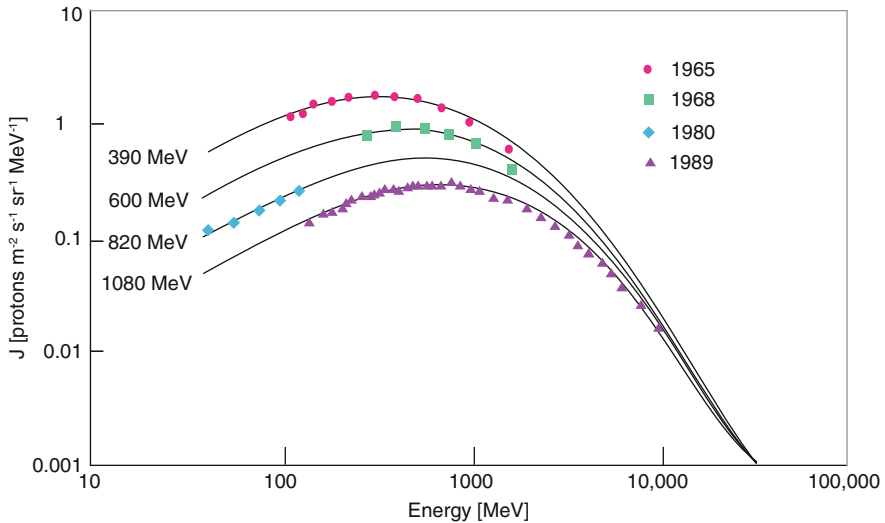


Fig. 5.7.3-1 Cosmic ray spectra observed from high-altitude balloons and satellites at four different times during 1965, 1968, 1980, and 1989. The *curves* are the best-fit spectra at Earth for modulation functions of (from the top) 390, 600, 820, and 1,080 MeV. After Bonino et al. (2001)

decelerated so that each cosmic ray has lost Φ MeV on its way through the heliosphere and arrives at Earth with energy E . Figure 5.7.2-1 illustrates this concept. Consider the application of (5.7.3-1a) to estimate the proton intensity at Earth, $J_E(E, \Phi)$, for a proton kinetic energy (at Earth) of 1,000 MeV and a modulation function $\Phi = 1,200$ MeV. The multiplicative term in (5.7.3-1a) is then $[1,000 \times (1,000 + 2 \times 938)] / [(1,000 + 1,200) \times (1,000 + 1,200 + 2 \times 938)]$ yielding $J_E(E, \Phi) = 0.321 J_{\text{LIS}}(E + \Phi)$. The corresponding Castagnoli and Lal local interstellar intensity, $J_{\text{LIS}}(E + \Phi)$, as depicted in Fig. 5.7.2-1 is 3.16 protons $\text{m}^{-2} \text{s}^{-1} \text{MeV}^{-1}$, and multiplication by 0.321 yields the intensity at Earth, 1.03 protons $\text{m}^{-2} \text{s}^{-1} \text{MeV}^{-1}$. While this seems complicated, modern calculators or computers make application of (5.7.3-1a) very straightforward.

Despite its approximations, experience has shown that the modulation function provides a remarkably good description of the cosmic ray spectrum observed at Earth. For example, Fig. 5.7.3-1 displays some accurate measurements of the cosmic ray spectrum made using instruments flown on high-altitude balloons and in satellites. It can be seen that each set of measurements (i.e., taken at different times during the solar cycle) is well described by (5.7.3-1), for a given value of the modulation function, Φ . In this use, it is essentially an empirical quantity, that is, the value of Φ is determined that provides the best fit to the observed data.

The original derivation of the modulation function, Φ , by Gleeson and Axford (1967, 1968) went on to show that it can be written in terms of the speed of the solar wind, V , and a scalar cosmic ray diffusion coefficient, $k(r)$, thus (Gleeson and Axford 1967; 1968):

$$\Phi = Ze \int \frac{V}{3k(r)} dr \quad (5.7.3-2)$$

where the integral is from the orbit of Earth to infinity. Here, the scalar diffusion coefficient is related to the mean free path (λ) of the motion of the cosmic ray by the equation

$$k(r) = \frac{v\lambda}{3} \quad (5.7.3-3)$$

where λ is the diffusion mean free path for the cosmic rays in the heliosphere, and v is the speed of the cosmic ray.

These two equations are a great assistance in understanding the manner in which the various heliospheric parameters influence the modulation. From (5.7.3-2), we see that Φ increases linearly as a function of the solar wind velocity, V (i.e., Φ doubles for a doubling of V). From (5.7.3-3), we note that a decrease in the diffusion mean free path leads to a decrease in $k(r)$, and from (5.7.3-2) that leads to an increase in Φ as well. That is, the modulation intensifies as the diffusive properties of the heliosphere increase (smaller λ and more frequent collisions with irregularities in the solar wind) and also as the speed of the solar wind increases. In this way, the modulation function provides an intuitive description of the manner in which the properties of the heliosphere affect the modulation seen at Earth. Figure 5.7.3-1 shows that the spectrum at Earth shifts downwards as Φ increases; that is, the modulation becomes more effective (“stronger modulation”).

Gleeson and Axford then introduced another quantity, the modulation potential (or parameter), defined by

$$\varphi = \int \frac{v}{3k(r)} dr \quad (5.7.3-4)$$

That is, the modulation function, Φ , and the modulation potential, φ , are related by

$$\Phi = Ze\varphi. \quad (5.7.3-5)$$

From this equation, we note that φ has the units of energy per unit charge, that is, the units of electrical potential. In practice, φ is usually measured in MV. Note further that the modulation potential contains all the heliospheric parameters and that, for a given observed degree of modulation, the value of φ is the same for protons, alphas, and heavies.

Having got that far, however, the cosmic ray research community then confounds everyone (including themselves) by confusing the concepts of the modulation function and potential. The units get confused; for example, modulation functions for use in (5.7.3-1) are sometimes quoted in MV (and modulation

Table 5.7.3-1 Defining the names, symbols, and units of the modulation function and modulation potential (parameter) as originally defined by Gleeson and Axford (1968), and as used throughout this book

Name	Symbol	Units	Dimension
Modulation Function	Φ	MeV	Energy
Modulation Potential (or parameter)	φ	MV	Electrical potential

Note: Some authors in the literature call φ the “force field parameter” or the “modulation strength”

potentials given in MeV). The fact that Gleeson and Axford used capital and lower-case phi, Φ and φ , for the two different quantities is also a major cause of confusion.

In an attempt to reduce this confusion, Table 5.7.3-1 defines the names, symbols, and units of these parameters.

We will use the modulation function in this book. The only good news about this whole confusing subject is that for cosmic ray protons (the majority of the cosmic rays), the numerical value of the modulation function (Φ) in MeV is the same as the numerical value of the modulation potential (φ) in MV. That is, for protons, when the modulation function in the modulation equation (5.7.3-1) is 1,000 MeV, the modulation potential is 1,000 MV.

The situation is more complicated for the alpha and heavier cosmic rays that constitute ~12% of the cosmic ray particles but can contribute >40% of the cosmogenic radionuclides that we observe (McCracken 2004). From (5.7.3-5), we can see that the modulation functions of alphas and protons are related by $\Phi_\alpha = 2\Phi_p$ and for heavier cosmic rays, $\Phi_h = Z\Phi_p$. As described in Sect. 5.8.2, the geomagnetic effects of protons, alphas, and heavies are different as well. Throughout this book, we will use the proton modulation function to quantify the degree of modulation of the galactic cosmic radiation. Note however that the alpha and heavier particles have been included in all the calculations of the production of the cosmogenic radionuclides, with due allowance for these differences. For many applications, it is not necessary to understand these complications, and they are therefore provided in the accompanying box. Anyone wishing to use Eq. (5.7.3-1) to compute cosmic ray fluxes or cosmogenic production from scratch must take the alpha and heavies into account and must therefore understand the content of the box.

5.7.3.1 BOX Modulation of the Alpha and Heavier Components of the Cosmic Radiation

From (5.7.3-5), it can be shown the modulation function of a cosmic ray of charge Z is given by

$$\Phi_Z = Z\Phi_p \tag{B5.7.3.1-1}$$

Using this and Eq. (5.7.3-1), it can be shown that the modulated spectrum for a cosmic ray with total energy per nucleon T_n , a charge Z , and atomic mass number A ,

(continued)

$$J_T(T_n, \Phi_p) = J_{\text{LIS}} \left(T_n + \frac{Z}{A} \Phi_p \right) \frac{T_n^2 - E_0^2}{\left(T_n + \frac{Z}{A} \Phi_p \right)^2 - E_0^2} \quad (\text{B5.7.3.1-2})$$

Note that this is identical to (5.7.3-1), except that the modulation function always appears as $(Z/A)\Phi_p$.

Equation (B5.7.3.1.2) shows that the flux of a specified component of the cosmic radiation at Earth, for a given total energy per nucleon of T_n , is given by the product of the LIS flux for that component at a total energy per nucleon which is $(Z/A) \Phi_p$ MeV higher, and another function of total energy per nucleon and $(Z/A) \Phi_p$ alone. It can be understood in general terms as if the cosmic ray LIS spectrum at energy per nucleon $(T_n + (Z/A) \Phi_p)$ had been decelerated so that each cosmic ray has lost $(Z/A) \Phi_p$ MeV on its way through the heliosphere and arrives at Earth with total energy per nucleon of T_n .

To illustrate the differences in the modulation of the protons and alpha cosmic rays, Fig. 5.7.3-2 plots the modulated spectrum for the proton modulation function $\Phi_p = 1,000$ MeV.

The geomagnetic field also results in substantial differences between the flux and nature of the protons, alpha, and heavier cosmic rays reaching the atmosphere. Anticipating results from Sects. 5.8 and 6.2, the production rate of either the cosmogenic radionuclides, or the counting rate of a neutron monitor at a geomagnetic latitude of λ , for a proton modulation function Φ_p , is approximated by the sum of the contributions made by the protons, alphas, and heavies, as given in (B5.7.3.1.2). Here, the contribution by the i th component is the integral from the “vertical cut-off energy” for the i th component, $E_{c,i}$, to infinity of the product of the modulated spectrum, $J_i(E, (Z/A)\Phi_p)$, multiplied by the specific yield function, $S_i(E)$ (see Sect. 6.5):

$$X(E_c, \Phi_p) = \sum_i \int_{E_{c,i}}^{\infty} S_i(E) J_i(E, (Z/A)\Phi_p) dE \quad (\text{B5.7.3.1-3})$$

Equation (5.8.2-2) gives the cut-off rigidity, P_c , of a point at geomagnetic latitude, λ . Using the relationship between rigidity and energy (Eq. 5.3-7b), the vertical cut-off energy of the i th component can be shown to be

$$E_{c,i} = \left[(Z/A)^2 P_c^2 + 0.938 \right]^{0.5} - 0.938 \quad (\text{B5.7.3.1-4})$$

Since $Z/A = 1$ for a proton, 0.5 for an alpha particle, and approximately 0.5 for heavies, it is clear that the vertical cut-off energy of a helium or heavy cosmic ray is approximately half that of a proton at low and medium geomagnetic latitudes.

As a consequence of (B5.7.3.1-2) and (B5.7.3.1-4), the shielding effects of the geomagnetic and heliomagnetic fields are only about half as effective for

$Z/A = 0.5$ compared to protons, and both effects mean that a greater proportion of the cosmogenic radionuclides are generated by helium and heavier cosmic rays, particularly at lower geomagnetic latitudes. The amplitudes of the modulation are different for the proton and helium components for the same reason.

Throughout this book, we quantify the degree of modulation of the observed cosmogenic radionuclides, or cosmic radiation, using the modulation function for the proton. Nevertheless, all calculations of the production of the cosmogenic radionuclides include the contributions made by the helium and heavier cosmic rays and are based on using (B5.7.3.1-2) to compute the modulated spectra of helium and heavier cosmic rays, Eq. (B5.7.1-4), to give the lower limits of integration in (B5.7.3.1-3). As noted elsewhere, the numerical value of the proton modulation function, Φ_p , is the same as that of the modulation potential, φ , which is the same for all components of the cosmic radiation.

The differences between the modulation and geomagnetic cut-off properties of the proton and helium components of the cosmic radiation are summarized in Fig. 5.7.3-2.

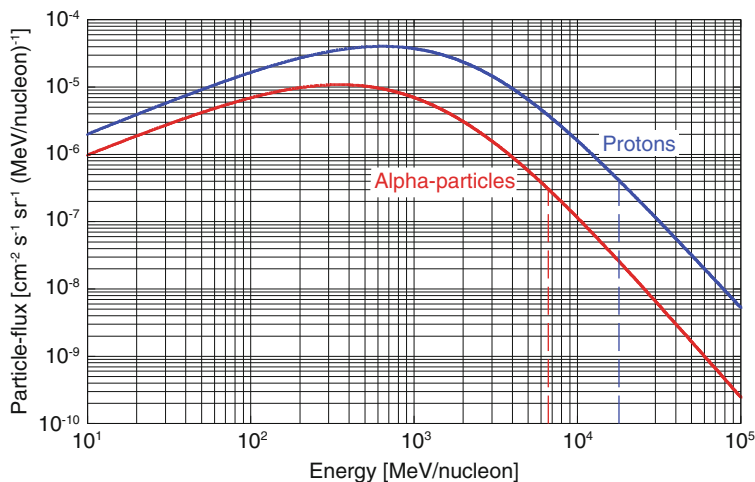
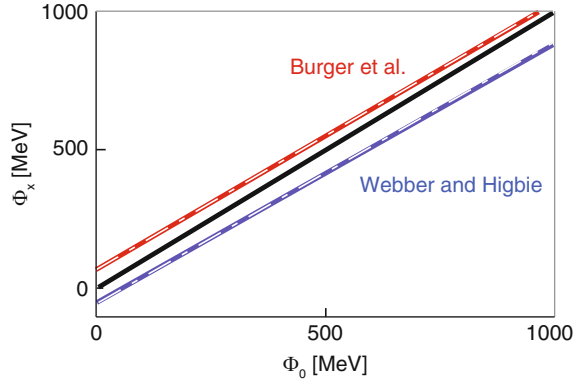


Fig. 5.7.3-2 Illustration of the differences between the proton and alpha particle cosmic radiation reaching the atmosphere near the equator using the Castagnoli and Lal LIS. There are two distinctly different effects, both due to the fact that for the same energy per nucleon, the alpha particles are less deflected than the protons and are therefore more able to penetrate both the heliomagnetic and geomagnetic fields. (1) The *two curves*, proton and helium quantify the difference in modulation by the heliomagnetic field for a proton modulation function of 1,000 MeV (modulation potential 1,000 MV in Eq. (B5.7.3.1-2)). (2) The *two vertical lines* indicate the lowest energy/nucleon that can reach the equator from the vertical [Eq. (B5.7.3.1-4)]

Fig. 5.7.3-3 Comparison of the modulation functions based on three formulations of the Local Interstellar Spectrum (LIS). After Steinhilber et al. (2008)



To compute the near-Earth spectrum for a given formulation of the LIS, Eq. (5.7.3-1) or (5.7.3-1a) is most commonly used. Usoskin and co-workers (Usoskin et al. 2005) and O'Brien and co-workers (O'Brien 1979), however, have used an alternative stochastic approach to compute the observed spectrum corresponding to any LIS and value of the modulation potential. The US Federal Aviation Authority (FAA) uses the O'Brien formulation in their advisory notifications of the radiation conditions at aircraft altitudes. As a consequence of this, and the several existing formulations of the LIS, there is no unique value of the modulation function (or potential) corresponding to an observed neutron monitor counting rate or cosmogenic measurement. This multiplicity may increase as new formulations of the LIS are developed in the future, as outlined in Sect. 5.7.2.

Using (5.7.3-1a), it is a straightforward matter to compute conversion relationships between the modulation functions based on different formulations of the LIS. Figure 5.7.3-3 illustrates this for the three LIS formulations discussed in the previous section. The conversion relationships can be written (Steinhilber et al. 2008):

$$\Phi_{\text{CL80}} = 1.04\Phi_{\text{B00}} - 73 \text{ MeV} \quad (5.7.3-6)$$

$$\Phi_{\text{CL80}} = 1.05\Phi_{\text{WH03}} + 54 \text{ MeV} \quad (5.7.3-7)$$

where Φ_{CL80} , Φ_{B00} , and Φ_{WH03} refer to the modulation functions based on the LIS given by Castagnoli and Lal, Burger et al., and Webber and Higbie, respectively.

For example, based on the Castagnoli and Lal spectrum, the cosmic ray intensities at the sunspot minima since 1954 are describable in terms of a modulation function $\Phi_{\text{CL80}} \sim 500$ MeV. Exactly the same cosmic ray intensities are quantified by $\Phi_{\text{WH03}} = 421$ MeV for the Webber and Higbie LIS, and $\Phi_{\text{B00}} = 545$ MeV for the Burger et al. LIS formulation.

As discussed in the next section, this book uses the Castagnoli and Lal local interstellar spectra (Castagnoli and Lal 1980) and the Gleeson and Axford definitions (Gleeson and Axford 1968) of the modulation function. It is clear

from the above and Fig. 5.7.3-1 that modulation functions computed at different times using a given LIS provide a useful parametric way to compare the cosmic ray intensities at all energies. However, it is vital to recognize that modulation functions obtained using different LIS must not be compared directly, without conversion using Fig. 5.7.3-3 or its equivalent.

5.7.4 *Practical Applications of the Modulation Function*

In Sect. 5.7.2 we have discussed the existence of several estimates of the LIS and the certainty that more will be developed in the future. Then in Sect. 5.7.3 we have reviewed the two parameters, the modulation function and modulation potential, and the confusion that is caused. We have further indicated that there are several different ways to define and compute the modulation functions and potentials. While these differences are the inevitable outcome of a topic that is still under scientific development, it causes frustration and uncertainty on the part of those wishing to utilize the cosmogenic record as an input to applications in atmospheric, astronomical, hydrological, and other application.

For this reason, we now state the convention that we will use uniformly throughout this book.

The problem as outlined above is similar to one faced by the ^{14}C dating community. As outlined in Sect. 23.2.1, the estimation of dates using ^{14}C is critically dependent on having an accurate value of the ^{14}C half-life. Recognizing that the accepted value (at that time) of 5568 years was likely to be improved over time, a two-step dating process was adopted. The first step estimates the “radiocarbon age” using the original half-life of 5568 years, thereby guaranteeing comparability of the “radiocarbon ages,” worldwide. The second step computes the “calibrated age” from the “radiocarbon age” using a calibration curve that allows for revisions of the half-life (now taken as 5730 years), and other calibration data.

To avoid confusion, we will employ a similar two-step procedure in respect of the LIS and the modulation function throughout this book. Recognizing that virtually all of the studies of the cosmogenic radionuclides up to ~2003 AD used the Castagnoli and Lal spectrum, we have adopted that LIS as the basis for all of our discussions and figures. For similar reasons, we have adopted the modulation function as defined by Caballero-Lopez and Moraal (2004) and Gleeson and Axford (1968).

We use Φ_{CL80} to designate the modulation function for protons based on the Castagnoli and Lal spectra, where 80 indicates the year of publication (1980). We stress, Φ_{CL80} has the dimensions of energy and is measured in MeV. It can be used in (5.7.3-1a) to compute the proton spectrum corresponding to any value of Φ_{CL80} . Its relationships to solar wind speed and the scattering properties of the heliospheric magnetic fields are given by (5.7.3-2). Its relationship to the modulation potential (corresponding to the Castagnoli and Lal spectrum) is given by (5.7.3-5).

The second step in this procedure will occur once we have direct experimental measurements of the LIS from outside the heliosheath (Fig. 5.6-1). Thus, Voyagers 1

and 2 may provide this by 2020 AD. Using the directly measured LIS, a calibration curve between Φ_{CL80} and Φ_{LIS} will be calculated applying (5.7.3-1a). With that, it will be possible to convert all previously estimated Φ_{CL80} to the “calibrated” estimate, Φ_{LIS} .

We stress, the modulation function should be regarded as a quasi-empirical parameter. It is based on a simplification of the accurate “cosmic ray propagation equation” (Eq. 5.7.1-1), and as such it includes the dependence on solar wind speed and on heliospheric magnetic field strength and scattering properties. While there are a number of major simplifications and approximations in its derivation (see Sect. 5.7.3), experience has shown that it provides a good, single parameter description of the majority of the cosmic ray observations made in the “instrumental” era and in the study of the ^{10}Be , ^{14}C , and other cosmogenic data for the past 10,000 years.

Equation (5.7.3-1a) shows that the modulated spectrum at energy E is computed using $J_{\text{LIS}}(E + \Phi)$, that is, the value of the LIS at an energy Φ greater than the energy of interest, E . From Box 5.7.2.1, the Castagnoli and Lal LIS is $9.9 \times 10^8 [E + 780 \exp(2.5 \times 10^{-4} E)]^{-2.65}$. Usoskin et al. (2005) have pointed out that Castagnoli and Lal failed to replace E with $(E + \Phi)$ in the term within the exponential. This has a small effect at the energies of importance for the production of the cosmogenic radionuclides and is well within the measurement errors of the cosmogenic data. Nevertheless, it will be appropriate to compensate for this small error when converting to the calibrated Φ_{LIS} , as outlined in this section.

5.7.5 Drift Effects (qA Positive and qA Negative Effects)

As shown in Fig. 5.5-4, the large-scale heliospheric magnetic field is curved in the shape of an Archimedes spiral as a consequence of the rotation of the Sun, together with the convection of the solar magnetic field radially outwards by the solar plasma. For field lines with a radius of curvature vector of $\mathbf{r}$, a cosmic ray will “drift” in the direction that is perpendicular to both $\mathbf{r}$ and the magnetic vector, $\mathbf{B}$. That is, as in Fig. 5.5-4, the drift is perpendicular to the plane of the paper. The drift rate is proportional to $\sin^2 \theta / (\mathbf{rB})$, where θ is the angle between the cosmic ray velocity vector and the magnetic vector. In other words, the drift is fastest for large θ , and small $\mathbf{r}$ (strongly curved magnetic lines of force) and small $\mathbf{B}$. This (and other drifts due to gradients in the strength of the magnetic field) leads to the “drift” (third) term in the cosmic ray propagation equation (5.7.1-1). The direction of the curvature vector is fixed in space by the constant rotation of the Sun, so the direction of the curvature drift is completely determined by the vector direction of the magnetic field and the sign of the charge on the drifting particle (Jokipii et al. 1977). If the polarity of the field reverses, the drift velocity reverses also. A positive cosmic ray (e.g., proton) drifts in the opposite direction to a cosmic ray electron.

Figure 5.7.5-1 displays the neutron monitor data observed since 1951. Note that the nature of the cosmic ray maxima (concurrent with the sunspot minima) alternates between relatively broad (1954, 1976, and 1996) and quite sharp (1965 and 1986). The same succession of broad- and sharp sunspot minima effects is seen in the satellite data

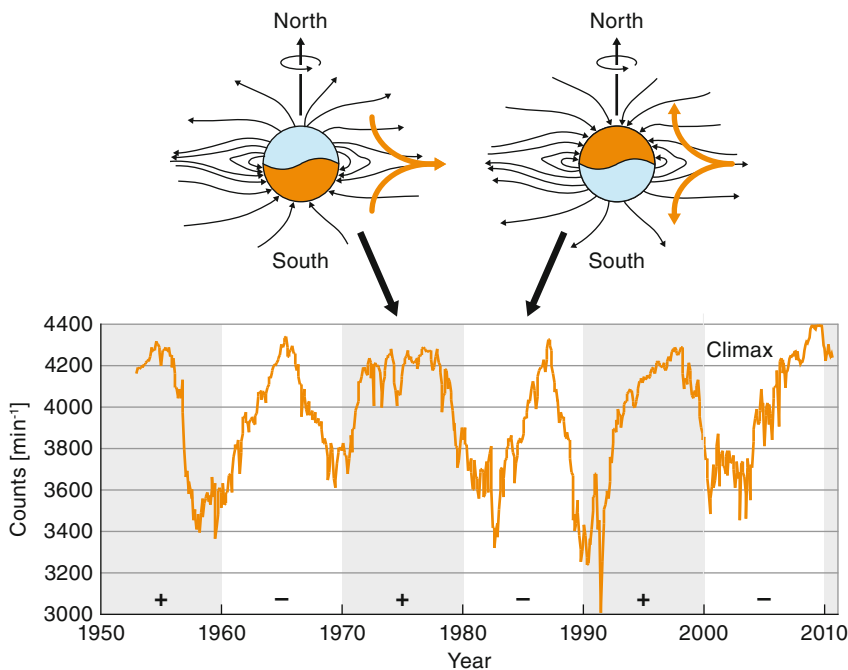


Fig. 5.7.5-1 The *bottom panel* presents the monthly Climax neutron monitor record from 1951 to 2007 and the normalized record from Hermanus (South Africa) for 2008–2010. The *plus* and *minus* signs indicate the polarity of the Sun’s polar field, as discussed in the text. The *cartoons* illustrate the combined effects of the drift motions experienced by cosmic rays in addition to the diffusive motion outlined in Sect. 5.7.1. During the positive half of the Hale magnetic cycle, some of the cosmic rays entering the heliosphere over the poles drift to the vicinity of the ecliptic and then leave the solar system along the “Parker spiral” field. During the negative half cycle, the drift motions are reversed

(see Fig. 7.2-2) This behaviour is a direct consequence of the drift effects and the 22-year “Hale” periodicity in the Sun’s magnetic fields, as outlined below.

The Sun’s polar field reverses near the maximum of the sunspot cycle. Its magnitude is sometimes represented by the symbol A , which is positive when the solar magnetic dipole is pointed into the northern sky. Referring to Fig. 5.7.5-1, it will be seen that the dipole pointed north until ~1958 and that it then reversed direction to point south until ~1969, when it reverted to the north orientation, and so on every 11 years.

Figure 5.7.5-1 illustrates the manner in which the drift effects influence the intensity of cosmic rays in the inner heliosphere. To allow for the influence of both the sign of the charge and the polarity of the magnetic field, we usually use the product qA to specify the direction of the drift effects, where q is the charge of the cosmic ray. The top panel in Fig. 5.7.5-1 shows that for $qA > 0$, the drift effects cause the cosmic rays to travel across the field lines from the polar regions of the Sun towards the ecliptic plane. That is, for the broad cosmic ray maxima with

$qA > 0$, some of the cosmic rays that spiral along the high-latitude magnetic lines of force reach the vicinity of Earth and hasten the recovery of the intensity from the minimum value attained near sunspot maximum.

Now consider the $qA < 0$ case in Fig. 5.7.5-1. The drift effects are ejecting cosmic rays from the inner heliosphere outwards along the lines of force of the polar magnetic field. That is, the drift effects are opposing the recovery of the intensity from the low values near sunspot maximum, and the intensity takes several years longer to recover to the sunspot minimum value than in the $qA > 0$ case. For the solar cycle 1954–1965, then, the intensity was still recovering at the commencement of the next solar cycle in 1965, leading to the sharp nature of the intensity maximum.

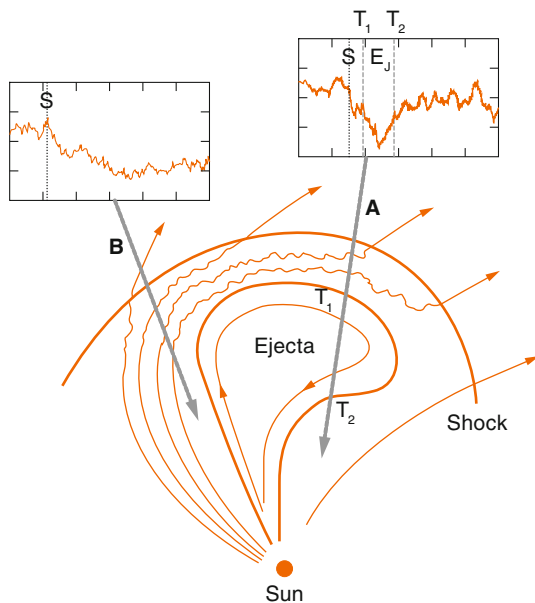
In summary, over the past 75 years (a very active Sun), the 22-year Hale cycle of the Sun's magnetic field has resulted in an alternating succession of broad and sharp cosmic ray maxima. The differences between the qA positive and negative cycles are most profound in the vicinity of the 7th and 8th year of the sunspot cycle; at that time, the intensity is still close to the minimum value for $qA < 0$, while it is close to fully recovered for $qA > 0$. For cosmic ray electrons, $qA < 0$ for the solar cycles with $qA > 0$ for protons and consequently, the broad–sharp sequence is out of phase with that for the positively charged cosmic rays. The drift effects also have a profound influence on the daily (diurnal) variation of the cosmic radiation at Earth and was the first evidence (1953) of the importance of the polarity of the solar magnetic fields upon the cosmic radiation; however, that has a negligible effect upon the production of the cosmogenic radionuclides and is beyond the scope of this book.

As Fig. 5.7.5-1 shows, the drift effects have an observable effect in the neutron and satellite data. This will be reflected in the production of the cosmogenic radionuclides and has led to studies of the ^{10}Be data to determine whether the polarity reversals of the solar magnetic field have suffered a phase change at the Grand Minima in the past. The results were inconclusive; however, the recent acquisition of more annual data series (from new ice cores) should allow the broad–sharp character to be determined for solar cycles throughout the past 700 years (see Chap. 17). Thus the cosmogenic data have the potential to provide important information about the past behaviour of the solar magnetic fields and the dynamo that generates them.

5.7.6 Shock Wave Effects (*The Forbush Decrease and GMIRs*)

As discussed in Sect. 5.5, a CME may eject solar plasma at a considerably higher speed than that of the ambient solar wind. This generates a shock wave that reaches the termination shock a year or more later. Figure 5.7.6-1 illustrates the magnetic fields associated with such a shock wave; note in particular that the magnetic field strength may be four times the ambient value in the region behind the shock. As a consequence, many cosmic rays are deflected and are unable to enter the regions within and behind the shock. The volume of the space behind the shock is steadily increasing as the shock moves outwards and consequently, both the density and the

Fig. 5.7.6-1 Illustration of the effect of an interplanetary shock wave on the magnetic fields in the heliosphere (Cane 2000). The magnetic field strength may be four times higher in the shocked region than in the ambient field. The most energetic shocks move outwards with speeds of $\sim 2000 \text{ km s}^{-1}$ compared to the $\sim 500 \text{ km s}^{-1}$ solar wind. Positions A and B illustrate the effects when the CME is launched from near the centre of the solar disc (A), and from 45° to the east of the central meridian (B)



intensity of cosmic radiation within and behind the shock is less than outside the region swept out by the shock.

Note that the nature of the shock wave effects depends strongly on the location of the Earth relative to the shock wave. For position A in Fig. 5.7.6-1, the most strongly shocked region envelopes the Earth, and the greatest depression of the cosmic radiation is observed soon after by Earth-based detectors. (This corresponds to the CME being launched from near the central meridian of the Sun).

Figure 7.4-1 presents the data from a very large Forbush decrease that approximates the situation depicted in model A. This shows that there were two distinct “steps” in the decreasing phase of the event: the first corresponding to the passage of the front surface of the shock wave past the Earth and the second to the arrival of the closed magnetic fields within the CME ejecta. The cosmic rays have to diffuse across the closed fields within the ejecta and, together with its expansion, this results in the large second step evident in the figure. In addition, the cosmic ray intensity is frequently depressed from the direction along the Parker field (from the direction of the Sun) for 6–12 h prior to the arrival of the shock wave at Earth. The event in Fig. 7.4-1 is discussed in more detail in Sect. 7.4.

For position B, the strongly shocked region has missed the Earth, and the region of depressed cosmic radiation only makes a glancing blow on the Earth. That is, the cosmic radiation decreases slowly and by a lesser amount than in case A. In this case, the CME was launched from about 45° to the east of the central solar meridian. We will discuss this mechanism again when we discuss the Forbush decrease and the Globally Merged Interaction Region (GMIR) in Sect. 7.4.

5.8 Geomagnetic Field Effects

The magnetic field of the earth (the “geomagnetic field”) is approximately 10^4 times stronger than the interplanetary magnetic field. It easily deflects the 1–20 GeV/nucleon cosmic rays recorded by satellite detectors and neutron monitors and which generate the majority of the cosmogenic radionuclides observed on Earth. Further, its strength has varied significantly over historic time. This, in addition to the modulating effects of the Sun, introduces major long-term changes into the intensity of the galactic cosmic radiation and the production rate of the cosmogenic radionuclides. The cosmogenic community refers to this as the “geomagnetic modulation” of the cosmic ray intensity. The geomagnetic signal is large in the cosmogenic record, and this has two consequences: (1) it provides an important means to investigate the manner in which the geomagnetic field has varied in the past; and (2) it can seriously interfere with the use of the cosmogenic data for other purposes. For both reasons, we now discuss the geomagnetic modulation effects in some detail.

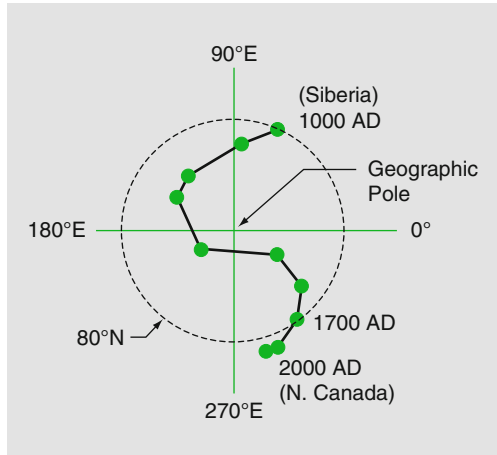
5.8.1 *The Properties of the Geomagnetic Field*

The geomagnetic field of the Earth approximates a classical magnetic dipole, with the “north-seeking” lines of force emerging from the southern hemisphere and re-entering the northern hemisphere. Over the past 200 years, the magnetic axis defined by the two geomagnetic poles has been offset from the centre of the Earth by ~440 km and inclined $\sim 11^\circ$ from the rotation axis. Presently, the north “dip” pole is located in northern Canada and the south geomagnetic pole off the coast of Antarctica. Note that they are not antipodal points – this is primarily due to the offset of the dipole from the centre of the Earth mentioned above.

The magnetic compass became an important tool for navigation in the fourteenth century. At first, it was thought to point to true north (i.e., to the rotation axis of the Earth), but it was later found to deviate by an amount that depended upon location on the Earth. It was soon recognized that the field approximated that of a piece of magnetite – that is a natural “bar magnet” with a dipole magnetic field. The exact position of the northern “dip” pole (where the magnetic field is vertical) was first determined by James Clark Ross in 1829, while seeking the “northwest” passage to the Orient. In 1904, Roald Amundsen located it again and found that it had moved northward from Ross’ observation. More recent observations have shown that it is moving quickly now towards the north. Looking back into the past, we find that the magnetic pole was in Siberia about 1000 AD, and it moved to its present position in northern Canada between the fourteenth and fifteenth centuries (Fig. 5.8.1-1). We will discuss the reason for this “polar wander” below.

Before we proceed, we mention a matter that can generate confusion among the wider scientific community. This is that (by definition) the terms “geomagnetic pole”

Fig. 5.8.1-1 The motion of the magnetic pole for the past 1,000 years (McElhinney and McFadden 2000). Note that the magnetic pole was in Siberia until ~1400 AD, and that it then moved rapidly to northern Canada



and “magnetic pole” do not mean the same thing. They can be hundreds of kilometres apart on the surface of the Earth, and our present knowledge indicates that the separation will vary with time. The position of the “geomagnetic pole” [derived from (5.8.1-1)] is the more important for the cosmic rays that generate the cosmogenic radionuclides, while it is the “magnetic pole” (i.e., the dip pole) that we learn about in our history books and have mentioned in the previous paragraph.

As discussed below, geology and archaeology allow us to determine some of the characteristics of the geomagnetic field in the past. Those measurements usually yield a “virtual magnetic pole,” which approximates the magnetic pole. Figure 5.8.1-1 shows that the position of the virtual magnetic pole (i.e., the magnetic pole) has moved from Siberia to Northern Canada over the past 1,000 years. As a consequence, the regions of the highest production of the cosmogenic radionuclides have been moving around as a function of time.

The geomagnetic field is primarily caused by electrical currents driven by convection of the viscous material in the “core” of the Earth (with a small contribution produced by the ionosphere and the magnetosphere, as discussed later). As this electrical dynamo (similar to that in the Sun) moves and waxes and wanes over time, so the intensity of the field and the direction of the magnetic axis also vary with time. The field and its secular variation can be derived from a potential function V . Assuming that the Earth is a sphere, V is usually expanded as a series of surface spherical harmonics:

$$V = \frac{a}{\mu_0} \sum_{l=1}^{\infty} \sum_{m=0}^l \left(\frac{a}{r}\right)^{l+1} P_l^m(\cos \theta) (g_l^m \cos(m\Phi) + h_l^m \sin(m\Phi)) \quad (5.8.1-1)$$

where

a : radius of Earth

r : distance from the centre of the Earth

θ : colatitude (90° - latitude)

Φ : longitude

P_l^m : Schmidt quasi-normalized form of the associated Legendre function of degree l and order m

g_l^m and h_l^m : the so-called Gauss coefficients, some of which are tabulated back to 1600 AD.

The northern, eastern, and vertical (downwards) components of the magnetic induction are then given by

$$B_x = -\mu_0(dV/d\theta)/r$$

$$B_y = -\mu_0(dV/d\Phi)/(r \cos \theta)$$

$$B_z = -\mu_0(dV/dr)$$

The magnitude of the Gauss coefficients decreases with increasing degree, so the higher order terms are small compared to the dipole terms. In the case of $m = 0$, the geocentric axial dipole ($l = 1$), quadrupole ($l = 2$), and octupole ($l = 3$) are described. Note that the terms in V decrease with the distance from the centre of the Earth (r) as the reciprocal of the $(l + 1)$ power of r . As a consequence, the dipole component dominates the internal field of the Earth at larger distances from the Earth. Nevertheless, as we will see in the next section, the Gauss coefficients up to the eighth order are now used to explain the observed distribution of the cosmic radiation on the surface of the Earth. The cosmogenic radionuclides are well mixed in longitude by atmospheric effects prior to sequestration in polar ice or trees, and in this case the effects of the higher order terms are largely averaged out. As a consequence, the dipole term largely determines the average production rate of the cosmogenic radionuclides as a function of latitude. However, this is not so for cosmogenic isotopes produced on the surface of the Earth (Chap. 11).

While direct measurements of the geomagnetic field only date back to the sixteenth century (for example, based on the >50,000 records taken by sailors when they “fixed their positions” all over the world at local noon in the seventeenth and eighteenth centuries), there are other ways to determine the nature of the geomagnetic field in the past. Two such methods are called “archeomagnetic studies” and “paleomagnetic studies” (McElhinney and McFadden 2000). The information they provide is important in the analysis of cosmogenic data, so we outline both of them here. Both depend on the fact that, at some time in the past, the magnetic domains in magnetite (a magnetic oxide of iron) in rocks have become aligned with the local direction of the geomagnetic field. Furthermore, frequently the degree of magnetization of the rocks was proportional to the strength of the geomagnetic field at that point on the Earth. By measuring the direction and the strength of the magnetization, then, it is possible to determine both the direction and the strength of the geomagnetic field in the past. If there is an independent way to

determine how long ago the magnetism was “locked” into the rocks, it is then possible to reconstruct the time dependence of the geomagnetic field in the past.

In the case of “archeomagnetism,” our ancestors started the measurement process for us when they lit a fire to cook their dinner or used clay to make a storage pot. In both cases, the magnetization of the rocks in the fireplace, or in the clay, was reset to the direction and strength of the local magnetic field by their actions. In this case, radiocarbon dating (e.g., from the carbon from the fire they lit) is frequently used to determine when our ancestors started the measurement for us. Since this method depends on human intervention at some earlier time, it is largely limited to investigations of the past 10,000 years. Using the measured direction of the magnetism, the “virtual magnetic pole” is determined, yielding a “polar wander” curve such as is given in Fig. 5.8.1-1. The degree of magnetization of the samples allows the strength of the geomagnetic dipole to be determined as a function of time, as shown in Fig. 5.8.1-2.

In the case of paleomagnetic measurements, it is the natural geological processes that initiated our measurements for us (McElhinney and McFadden 2000). When volcanic lava cooled through the “Curie point” ($\sim 500^\circ\text{C}$), its magnetic properties were reset to the strength and direction of the local magnetic field. Alternatively, as sediment accumulated at the bottoms of lakes or the oceans, the magnetite grains in the sediment lined up with the local magnetic field and were later locked in place when the sediment reached a certain degree of compaction. In both cases, measurement of the magnetic properties of the rocks allows us to determine some of the properties of the geomagnetic field over $>1,000,000$ years into the past.

Figures 5.8.1-1 and 5.8.1-2 show that the Earth’s magnetic field has changed greatly over a very small interval of geological time. Firstly, Fig. 5.8.1-1 shows that

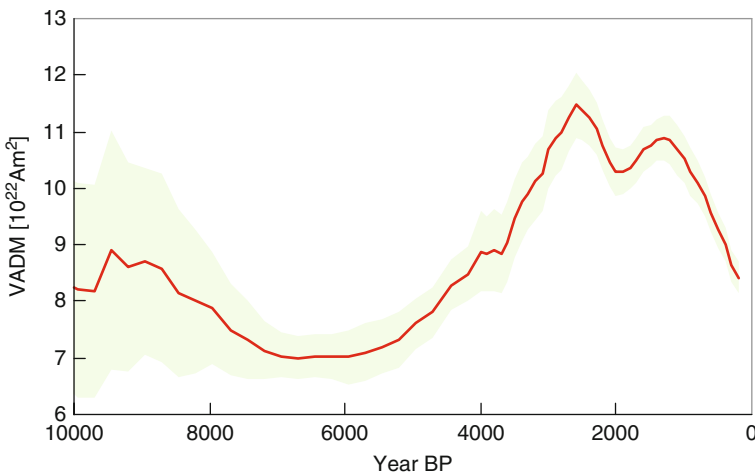


Fig. 5.8.1-2 The strength of the geomagnetic dipole over the past 10,000 years derived from archeomagnetic data and expressed as the virtual axial dipole moment (VADM). The present-day value is $8.0 \times 10^{22} \text{ Am}^2$ (Knudsen et al. 2008). The band reflects the estimated 2σ -errors

the vector direction of the geomagnetic dipole has changed by a surprising degree; some studies suggest that the dipole may have made an angle of up to 20° with the rotation axis at times in the past. Further, Fig. 5.8.1-2 shows that the strength of the Earth's field has changed by a factor of approximately 1.7 over the past 5,000 years. The paleomagnetic method also yields a most remarkable result: that the Earth's magnetic field has reversed its polarity in the past. Further, sometimes, but not always in association with magnetic reversals, the strength of the magnetic dipole decreased to a low value ($\sim 10\%$) compared to the present-day value. All these changes resulted in large changes in the rate of production of the cosmogenic radionuclides in the past (Chaps. 10 and 21).

Finally, it should be commented that there are many sources of error in both archaeomagnetic and paleomagnetic studies. They may be of human origin – human agency may have moved the rocks or pots on which the measurements are based. Earthquakes, subsidence, or other geological processes may have moved the sedimentary or volcanic rocks. Lightning strikes or chemical reactions may have remagnetized the rocks or pots in the intervening years. In addition, it may be difficult to determine how long ago the magnetism was locked in place. These several uncertainties can be reduced by averaging over a large number of samples, obtained at widely separated locations. In addition, both techniques are steadily improving in their ability to quantify the properties of the historic geomagnetic field, and each improvement will yield a direct benefit to the study and use of the cosmogenic radionuclides.

5.8.2 *The Geomagnetic Cut-off Rigidity*

The 1–100 GeV cosmic rays that produce the cosmogenic radionuclides and are recorded by neutron monitors and satellite detectors are strongly deflected by the geomagnetic field.

As discussed in Sect. 5.4, the cosmic rays reaching Earth have different values of Z (atomic number) and A (atomic mass number). For the same energy per nucleon, a ^4He cosmic ray has four times the momentum of a proton of the same energy. It only has twice the electrical charge of the proton, so the magnetic forces on the ^4He cosmic ray are only twice those on the proton (see Sect. 5.3). Therefore, ^4He is less accelerated (i.e., less deflected) from its path by the magnetic field than the proton (Newton's second law of motion), and it may penetrate the geomagnetic field more easily, and gain access to parts of the Earth that a proton of the same energy per nucleon cannot reach. To simplify the problem, it is the usual convention to discuss the geomagnetic cut-off and other magnetic effects in terms of the magnetic rigidity (= momentum to charge ratio) of the cosmic rays (see Sect. 5.3). Since about 40% of the cosmogenic radionuclides are produced by the alpha particle (He) and heavier components of the cosmic radiation, it is particularly important that we take the differences between their behaviour, and that of protons, into account.

While we have chosen to use energy per nucleon (E) in the rest of this book in preference to rigidity, P , we make an exception in this discussion of the geomagnetic effects. The reasons are several: it is the usual convention to do so, the equations are simpler, and it is consistently used for this purpose in the literature.

Equation (5.3.7) and Fig. 5.3-1 describe the relationship between rigidity and kinetic energy per nucleon. We recapitulate them here: $P = (A/Z) [(E_0 + E)^2 - (E_0)^2]^{0.5}$ and $E = [(Z/A)^2 P^2 + (E_0)^2]^{0.5} - E_0$. Here P is the rigidity in GV, and the cosmic ray has electric charge Z , atomic mass A , energy E GeV/nucleon, and E_0 rest energy/nucleon. For a proton, $(A/Z) = 1$ and ≈ 2 for all other nuclei. For the same energy per nucleon, the rigidity of a proton is half that any of the other cosmic ray nuclei reaching Earth. That is, as discussed above, its orbit is less “rigid” or more easily bent by the magnetic field. We shall see that this means that lower energy (per nucleon) helium and “heavy” nuclei can reach a point at the top of the Earth’s atmosphere than is the case for protons.

Figure 5.8.2-1 shows the orbits traced by a number of cosmic rays as they approach the Earth. The details of an orbit depend strongly upon the values of the Gauss coefficients (Sect. 5.8.1) that quantify the geomagnetic field, the particle rigidity, and where it ultimately intersects with the Earth, and from what direction. Luckily, for most practical purposes, this degree of detail is not warranted. We can then use an approximation called the “geomagnetic cut-off rigidity,” P_c that avoids

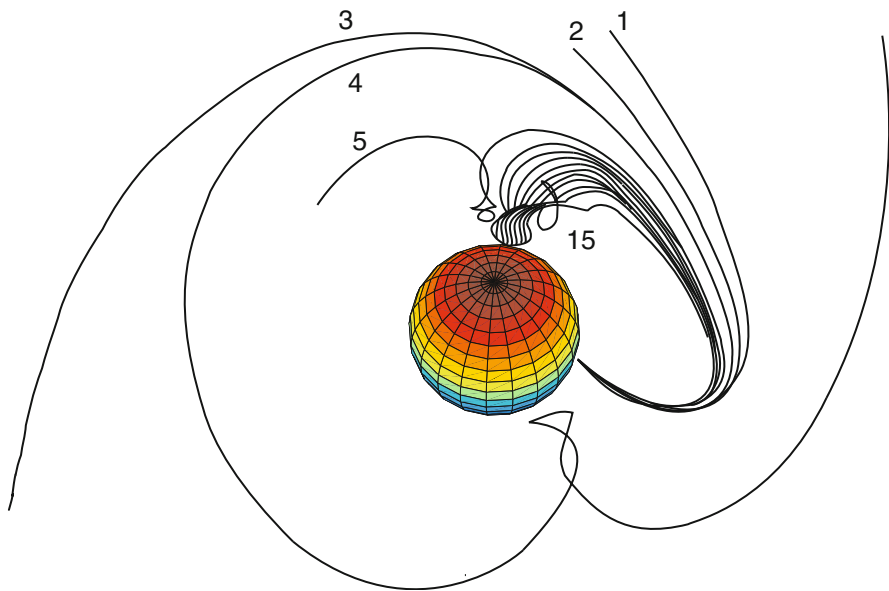


Fig. 5.8.2-1 Various paths taken by cosmic rays with different rigidities that ultimately arrive at Washington, DC, from the vertical direction. The highest rigidities (numbers 1, 2, and 3) follow relatively simple paths. Lower rigidities follow more complicated paths (e.g., numbers 4 and 15). Near the cut-off rigidity, the cosmic rays all follow complicated paths and may circle the Earth before impacting the atmosphere (Smart et al. 2000)

the complexities introduced by the more complex orbits such as those labelled 5–15 in Fig. 5.8.2-1. While there are some subtleties called the “shadow cones,” their effects are relatively small, and we define an effective cut-off rigidity P_c for any point on Earth and any direction of arrival. All rigidities less than P_c are said to be “forbidden,” while those above P_c are “allowed.” The cosmic ray intensity recorded by a neutron monitor, or which is available to produce cosmogenic radionuclides, is then approximated by the integral of the cosmic ray spectrum from P_c to infinity.

For many practical purposes, and for ease of understanding, we first restrict our treatment to the dipole component of the geomagnetic field. For a cosmic ray to reach the top of the Earth’s atmosphere at a certain geomagnetic latitude, λ , solution of the equations of motion shows that a minimum magnetic cut-off rigidity P_c is necessary. This “cut-off rigidity” is given by

$$P_c = pc = \frac{M}{R^2} \frac{\cos^4 \lambda}{\left[(1 + \cos \theta \cdot \cos^3 \lambda)^{1/2} + 1 \right]^2} \quad (5.8.2-1)$$

where M and R are the magnetic dipole moment and radius of the Earth, respectively. For the present day Earth, write $M = M_0$, then $M_0/R^2 \approx 59.6$ GV. The cut-off rigidity given by this equation is often called the “Stoermer cut-off,” in honour of the Swedish scientist who derived this equation at the beginning of the twentieth century in his studies of the aurora borealis.

Note that the cut-off rigidity depends upon the angle θ , the angle between the trajectory of the arriving particle and the vector to the (magnetic) West in the horizontal plane. That is, $\theta = 0^\circ$ for a cosmic ray arriving from the western horizon; $\theta = 90^\circ$ for any direction in the (magnetic) North–South vertical plane, and $\theta = 180^\circ$ for a cosmic ray arriving from the eastern horizon.

This equation provides an insight into a number of important properties of the cosmic radiation at the surface of Earth, and the production of cosmogenic radionuclides, as we now discuss.

(a) The Vertical Stoermer Cut-Off (Dipole Magnetic Field)

For cosmic rays arriving at the top of the atmosphere from the vertical direction, $\theta = 90^\circ$. From Eq. (5.8.2-1), the vertical cut-off rigidity, $P_{c,v}$, is then given by

$$P_{c,v} = 14.9 (M/M_0) \cos^4 \lambda \quad (5.8.2-2)$$

and is plotted in Fig. 5.8.2-2. This equation tells us that, at present, a cosmic ray with a rigidity < 14.9 GV cannot reach the geomagnetic equator from the vertical direction. From Eq. (5.3-7b) (or extrapolation of Fig. 5.3-1), this implies a cut-off energy of $E_c = 14$ GeV for protons and 6.5 GeV/nucleon for helium and heavier nuclei. Note that while the cut-off rigidity is the same for protons and alpha particles, the cut-off energies (in GeV per nucleon) are quite different.

(b) The Temporal Variability of the Stoermer Cut-Off

Fig. 5.8.2-2 The (Stoermer) vertical cut-off rigidity for the dipole component of the Earth's magnetic field during the instrumental era. Note that the maximum cut-off rigidity for $\lambda = 0$ has varied between 12.9 and 21.7 GV over the past 6,000 years (see discussion in text). When the non-dipole components of the geomagnetic field are taken into account, the variation has been even greater

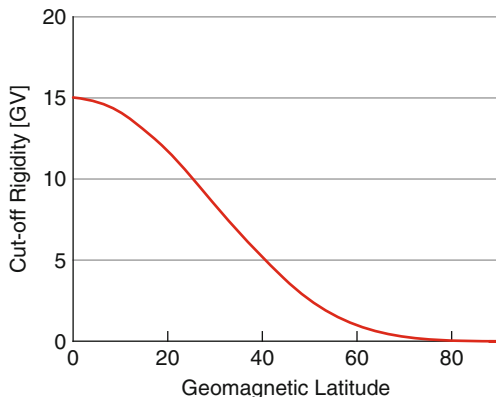


Figure 5.8.1-2 shows that the magnetic dipole moment of the Earth has varied from 87% to 146% of the present-day value over the past 10,000 years. Equation (5.8.2-2) shows that the equatorial vertical cut-off rigidity has varied from 12.9 to 21.7 GV, and this implies that the cosmic radiation intensity near the equator (and at all latitudes) has varied to a substantial degree over time. For example, the production rate of ^{10}Be at the equator (for $\Phi = 550$ MeV) has varied by a factor of approximately 1.4 over that time (see Fig. 10.3.3-3).

As illustrated in Fig. 5.8.1-1, the magnetic pole has exhibited “polar wander,” and this has a very substantial effect upon the vertical cut-off rigidity of any point on Earth. As an example, consider the vertical cut-off for the dipole approximation to the geomagnetic field. For the point at geographic coordinates (48°N , 285°E) near Ottawa, Canada, the Stoermer vertical cut-off rigidity was 1 GV in 2000 AD and 5.2 GV in 1000 AD when the north geomagnetic pole was in Siberia (Fig. 5.8.1-1) but without allowance for the changing magnetic moment of Earth. Allowing for that (from Fig. 5.8.1-2), the cut-off rigidity was higher still at 7.1 GV. That is, the cut-off rigidity at this one location has varied from 7.1 to 1.0 GV in the past 1,000 years. The calculations in Section 10.3 show that this resulted in a factor of 1.8 increase in ^{10}Be production (for $\Phi = 550$ MeV) near Ottawa over that interval. This example shows that the cosmic ray intensity at the surface of Earth, and the cosmogenic production rate, are subject to substantial “geomagnetic modulation.” This will be quantified in Sect. 10.3.

(c) The East–West Effect

For cosmic rays arriving at the equator, with $\theta = 0^\circ$, that is from the western horizon, Eq. (5.8.2-1) gives the cut-off rigidity = 10.3 GV. For $\theta = 180^\circ$, that is from the eastern horizon, the cut-off rigidity = 59.6 GV. That is, near the equator, many more cosmic rays arrive from the west than from the east. The inspection of (5.8.2-2) shows that this “east–west” effect decreases as θ approaches 90° and becomes smaller at higher geomagnetic latitudes (λ). Figure 5.8.2-3 plots Eq. (5.8.2-1) for all directions of arrival for all geomagnetic latitudes and shows that the cosmic ray intensity and cosmogenic production is far from isotropic in the equatorial regions of the Earth. For example, for $\lambda = 0^\circ$, the cut-off rigidities for

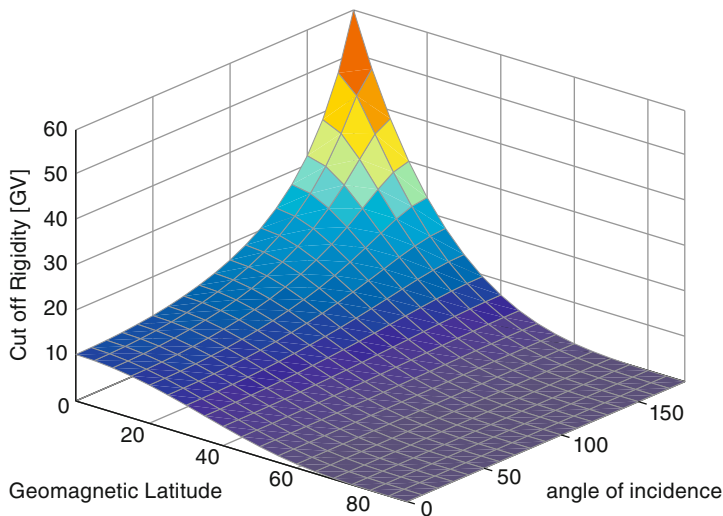


Fig. 5.8.2-3 A three-dimensional presentation of the cut-off rigidity given by Eq. (5.8.2-1)

the cone with a generating angle of 45° to the east are all >30 GV, while they are ≤ 12 GV for a similar cone from the west. Since the primary cosmic ray spectrum varies as $E^{-2.37}$ (Sect. 5.4), the cosmic ray intensity from within the western cone is a factor of ~ 8.8 times greater than that from the eastern cone.

The East–West effect may have significance for exposure dating. Thus the vertical cut-off rigidity may not be appropriate for exposure dating of say, steep east- and west-facing hills in the equatorial zone. The fact that polar wander may have changed the geomagnetic latitude of a sampling site by as much as 40° in the past may also require consideration.

(d) The Effects of the Non-dipole Terms

Equations (5.8.2-1) and (5.8.2-2) say that for a dipole field, with the dipole at the centre of the Earth, the cut-off rigidity is 14.9 GV at all points around the geomagnetic equator, and if true, this would mean that the rate of production of the cosmogenic radionuclides would be the same at all points around the geomagnetic equator as well. However, the manner in which the geomagnetic dipole is presently offset from the centre of the Earth by ~ 440 km, and other higher order terms in the expansion of the geomagnetic field [Eq. (5.8.1-1)], mean that this is not so. Further, as shown by Figs. 5.8.1-1 and 5.8.1-2, the position of the magnetic poles and the strength of the geomagnetic field have varied quite rapidly with time.

In the general case, the cut-off rigidities, P_c , can be calculated for any set of Gauss coefficients in (5.8.1-1). For any location on Earth, for a given direction of arrival, step-by-step integration of the relativistic equations of motion in the geomagnetic field determines whether that rigidity is allowed or forbidden. This is

repeated many times to determine the effective cut-off for that place and direction of arrival (Smart et al. 2000).

Figure 5.8.2-4 displays the contours of vertical cut-off rigidity computed in this manner for 1600 and 2000 AD, superimposed on a map of the world. In general, the computed cut-off rigidities for the present epoch are in good agreement with neutron monitor and other observations and can be used with confidence. In practice, they are frequently quite different from the value given by the simple dipole field calculations.

Note also the substantial changes in the cut-offs in the 400 years that elapsed between 1600 AD and 2000 AD. In the 1600 AD diagram, the highest cut-off rigidities (>17 GV) were in the vicinity of the Panama Canal. In the 2000 AD diagram, they are on the opposite side of the globe, having moved 160° west to the vicinity of Singapore. Note also that the vertical cut off for Cape Town has changed from ~ 12 GV in 1600 AD to ~ 4.5 GV in 2000 AD.

In passing, we note that Gauss coefficients up to the third degree [$\ell = 3$ in (5.8.1-1)] are available from 1600 AD to the present, and that cut-off maps similar to Fig. 5.8.2-4 have been published for 1700, 1800, and 1900 AD, as well. They show a systematic change from one century to the next (Shea and Smart 2004).

In summary, polar wander, the changing magnetic dipole moment, and changes in the higher order terms in the geomagnetic field mean that there have been quite large changes in the distribution of cosmic ray intensity on the Earth over a very short interval of time. It is only reasonable to assume that similar fast changes have occurred throughout the Holocene (since 10,000 years ago) and during geological history.

Clearly, the change in magnetic moment in the past will have left their trace in the cosmogenic record. Changes such as that for Cape Town, due to higher order changes in the geomagnetic field, may have discernible impact upon exposure dating (Chap. 11) and the injection profiles for the cosmogenic nuclide production used in conjunction with global circulation models (Chap. 13).

5.8.3 *The Earth's Magnetosphere and the Polar Aurora*

At distances larger than about 5 Earth radii, the symmetric dipolar magnetic field of the Earth is strongly modified by the interaction with the solar wind (Fig. 5.8.3-1). The solar magnetic field embedded in the solar wind, and the wind itself, compresses the geomagnetic field on the dayside forming a shock wave called the bow shock in analogy to the waves ahead of the bow of a ship. Like a rock in a fast-flowing river blocks the water and diverts it, at about 10 Earth radii the Earth's magnetosphere directs the solar wind around the Earth. At the bow shock, the solar wind abruptly decelerates, and the slower, hotter, solar plasma flows out to the magnetic tail. The magnetic tail of the Earth extends on the night side to a distance of up to 300 Earth radii ($\sim 2 \times 10^6$ km). From the magnetic tail, the plasma is injected into the "plasma sheet" and travels back towards the Earth. The circulation

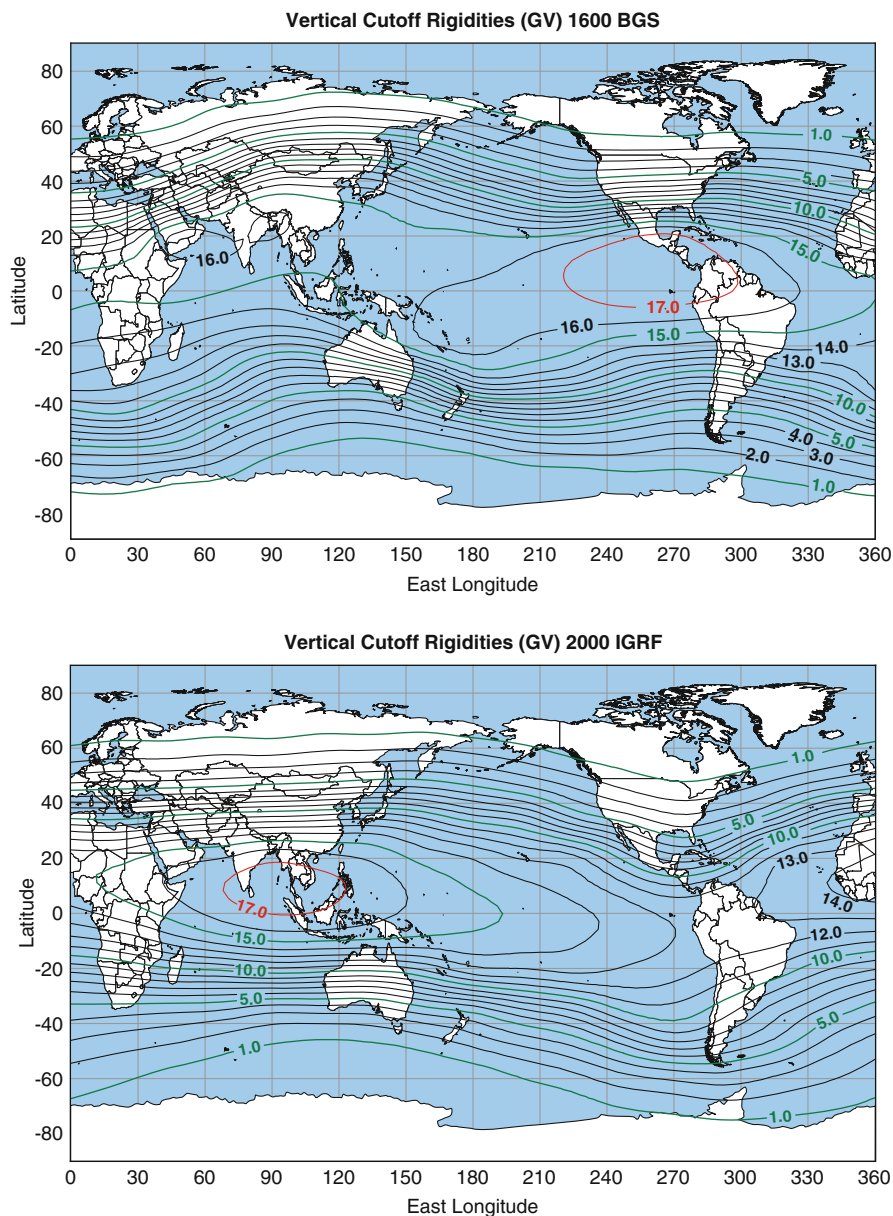


Fig. 5.8.2-4 The worldwide vertical cut-off rigidities (Smart and Shea 2009). Panel (a) for 1600 AD. Note the highest cut-off rigidities are in the vicinity of the Panama Canal. Panel (b) for 2000 AD. The region of highest cut-off has moved 160° west to the vicinity of Singapore. Comparison will show that the cut-off has changed by 4 GV for many parts of the world, with the change being >7 GV for some locations

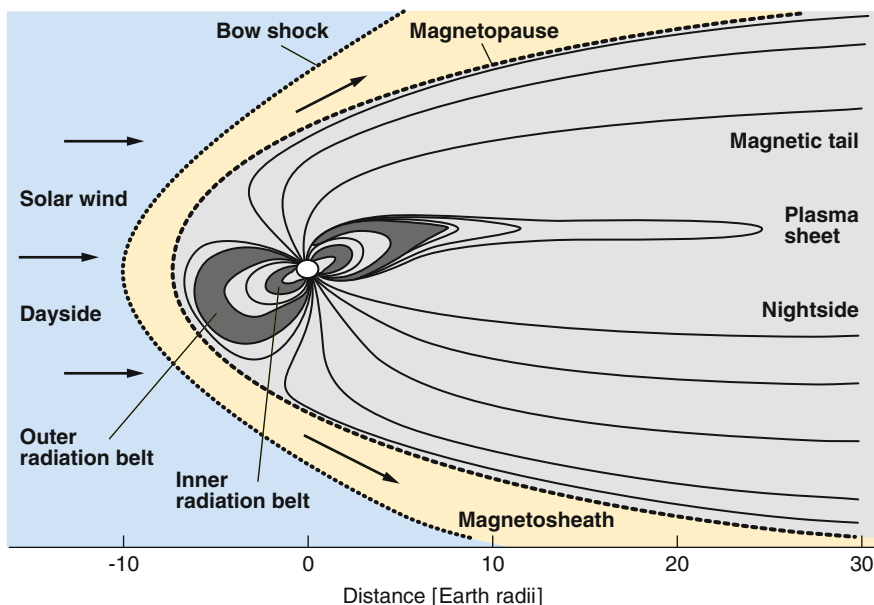


Fig. 5.8.3-1 The Earth's magnetosphere

of the plasma is driven by energy extracted from the solar wind. Merging of solar and terrestrial magnetic fields slows down the plasma in the vicinity of the magnetic tail.

The magnetosheath is a turbulent layer immediately behind the bow shock where particles (mainly protons and electrons) are accelerated to energies several thousand times as great as in the solar wind. The universality of nature is evident here; as in the shock waves associated with supernova; at the termination shock in the heliosphere; and in CME, we see that ions and electrons are accelerated to higher energies in the magnetospheric shock wave. The difference here is that the energies attained in the magnetospheric bow shock are smaller than those in the other examples.

The entire magnetosphere is very variable. Depending on the solar wind density and velocity, it compresses and expands and changes its size and shape constantly. These changes are particularly significant when an interplanetary shock wave (driven by a CME) arrives at Earth. This results in large fluctuations in the strength and direction of the magnetic field at the surface of Earth, called a "geomagnetic storm." The ever-present changes in the geomagnetic field do affect the cosmic ray cut-off rigidities to a small degree, particularly following a large geomagnetic storm. Then the cut-off rigidities at the top of the atmosphere may change by several GV and the production rate of the cosmogenic radionuclides would change to a small degree. However, these "storm-time" changes in the cut-off only last for several days at most and will not result in a measurable change in the annual production rate of the cosmogenic radionuclides.



Fig. 5.8.3-2 The aurora, an important source of information about past solar activity. *Left Panel:* the aurora borealis seen in Augsburg, Germany, on September 10, 1580. Painted by Bartholme Kaeppler. (ZB Graphische Sammlung, Zurich). *Right Panel:* a modern photograph of an aurora

The magnetosphere and the acceleration of particles associated with the bow shock and in the geomagnetic tail result in the *aurora borealis* in the northern hemisphere and the *aurora australis* in the southern. The accelerated electrons follow the geomagnetic lines of force into the polar atmosphere, where they ionize the atmospheric gasses. The ions then recombine, emitting visible light at the wavelengths characteristic of the atmospheric atoms. Most aurorae emit green light; in very large magnetic storms, red aurorae are produced. Large aurorae are quite spectacular and far surpass the best light shows made to entertain the masses using our modern technologies. Almost every night, aurora are observed in the “auroral ovals” that are at a “geomagnetic latitude” of 22° from the geomagnetic pole. However, during a large geomagnetic storm, they can be seen much closer to the equator; in recent history, a “great red” aurora was reported in Bombay, India (1859) and another in Indonesia (1805). In Roman times, the Roman guard was sent from Rome to the city of Ostia when a red aurora was mistaken for a large fire in that city.

In medieval times, the aurora caused wonder, and sometimes alarm, and were frequently recorded in the written records and illustrations of the time (Fig. 5.8.3-2). That use of that ubiquitous scientific instrument, the human eyeball, therefore means that we have quite extensive records of auroral activity since medieval times in Europe and for the past 2,000 years in the Orient. They tell us when the Sun was active, and this provides some of the key evidence that confirms that the majority of the short- term (decadal and century long) changes in the production rates in the cosmogenic record are a consequence of solar activity (Chap. 17, also Fig. 7.3-1).

References

- Aschwanden MJ (2005) Physics of the solar corona. An introduction with problems and solutions. Springer, New York
- Bonino G, Cini Castagnoli G, Cane D, Taricco C, and Bhandari N (2001) Solar modulation of the galactic cosmic ray spectra since the Maunder minimum. Paper presented at ICRC2001, Copernicus Gesellschaft 2001
- Burger RA, Potgieter MS, Heber B (2000) Rigidity dependence of cosmic ray proton latitudinal gradients measured by the Ulysses spacecraft: implications for the diffusion tensor. *J Geophys Res* 105(12):27447–27455
- Caballero-Lopez RA, Moraal H (2004) Limitations of the force field equation to describe cosmic ray modulation. *J Geophys Res-Space Phys* 109:A01101
- Caballero-Lopez RA, Moraal H, McCracken KG, McDonald FB (2004) The heliospheric magnetic field from 850 to 2000 AD inferred from Be-10 records. *J Geophys Res-Space Phys* 109:A12102
- Cane HV (2000) Coronal mass ejections and forbush decreases. *Space Sci Rev* 93(1–2):55–77
- Castagnoli G, Lal D (1980) Solar modulation effects in terrestrial production of C-14. *Radiocarbon* 22(2):133–158
- Eddy JA (1976) The Maunder minimum. *Science* 192(4245):1189–1201
- Eddy JA (1977) Case of missing sunspots. *Sci Am* 236(5):80–92
- Garcia-Munoz M, Mason GM, Simpson JA (1975) Anomalous He-4 component in cosmic-ray spectrum at less than equal to 50 Mev per nucleon during 1972–1974. *Astrophys J* 202(1):265–275
- Gleeson LJ, Axford WI (1967) Cosmic rays in the interplanetary medium. *Astrophys J* 149:L115–L118
- Gleeson LJ, Axford WI (1968) Solar modulation of galactic cosmic rays. *Astrophys J* 154:1011–1026
- Herbst K, Kopp A, Heber B, Steinhilber F, Fichtner H, Scherer K, Matthia D (2010) On the importance of the local interstellar spectrum for the solar modulation parameter. *J Geophys Rese-Atmos* 115:D00120
- Hoyt DV, Schatten KH (1998) Group sunspot numbers: a new solar activity reconstruction. *Solar Phys* 179:189–219
- Jokipii JR (1991) Variations of the cosmic-ray flux with time. In: Sonett CP, Giampapa HS, Mathews MS (eds) *The Sun in Time*. University of Arizona Press, Tucson, AZ, pp 205–220
- Jokipii JR, Levy EH, Hubbard WB (1977) Effects of particle drift on cosmic-ray transport. 1. General properties, application to solar modulation. *Astrophys J* 213(3):861–868
- Knudsen MF, Riisager P, Donadini F, Snowball I, Muscheler R, Korhonen K, Pesonen LJ (2008) Variations in the geomagnetic dipole moment during the Holocene and the past 50 kyr. *Earth Planet Sci Lett* 272(1–2):319–329
- Lal D, Peters B (1962) Cosmic ray produced isotopes and their application to problems in geophysics. In: Wilson JG, Wouthuysen SA (eds) *Progress in elementary particle and cosmic ray physics*. North-Holland publishing company, Amsterdam
- Langner UW, Potgieter MS, Webber WR (2003) Modulation of cosmic ray protons in the heliosheath. *J Geophys Res-Space Phys* 108:D08039
- Masarik J, Beer J (1999) Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res* 104(D10):12099–12111
- Masarik J, Beer J (2009) An updated simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res-Atmos* 114:D11103
- Masarik J, Reedy RC (1995) Terrestrial cosmogenic-nuclide production systematics calculated from numerical simulations. *Earth Planet Sci Lett* 136:381–395
- McCracken KG (2004) Geomagnetic and atmospheric effects upon the cosmogenic Be-10 observed in polar ice. *J Geophys Res-Space Phys* 109:A04101

- McDonald FB (1998) Cosmic-ray modulation in the heliosphere – a phenomenological study. *Space Sci Rev* 83(1–2):33–50
- McElhinney MW, McFadden PL (2000) Paleomagnetism: continents and oceans. Academic, San Diego, CA
- Moskalenko IV, Strong AW, Ormes JF, Potgieter MS (2002) Secondary antiprotons and propagation of cosmic rays in the galaxy and heliosphere. *Astrophys J* 565(1):280–296
- O’Brien K (1979) Secular variations in the production of cosmogenic isotopes in the Earth’s atmosphere. *J Geophys Res* 84:423–431
- Parker EN (1958) Dynamics of the interplanetary gas and magnetic fields. *Astrophys J* 128(3):664–676
- Parker EN (1963) Interplanetary dynamical processes. Wiley, New York
- Potgieter MS, Leroux JA (1992) The simulated features of heliospheric cosmic-ray modulation with a time-dependent drift model. 3. General energy-dependence. *Astrophys J* 392(1):300–309
- Potgieter MS, Moraal H (1985) A drift model for the modulation of galactic cosmic-rays. *Astrophys J* 294(2):425–440
- Schlickeiser R (2003) Astrophysics. Springer, Berlin
- Schwabe H (1844) Sonnen-Beobachtungen im Jahre 1843. *Astron Nachrichten* 495:233–236
- Shea MA, Smart DF (2004) Preliminary study of cosmic rays, geomagnetic field changes and possible climate changes. *Adv Space Res* 34(2):420–425
- Smart DF, Shea MA (2009) Fifty years of progress in geomagnetic cutoff rigidity determinations. *Adv Space Res* 44(10):1107–1123
- Smart DF, Shea MA, Fluckiger EO (2000) Magnetospheric models and trajectory computations. *Space Sci Rev* 93(1–2):305–333
- Steinhilber F, Abreu JA, Beer J (2008) Solar modulation during the Holocene. *Astrophys Space Sci Trans* 4:1–6
- Stuiver M, Quay PD (1980) Changes in atmospheric C-14 attributed to a variable Sun. *Science* 207(4426):11–19
- Swordy SP (2001) The energy spectra and anisotropies of cosmic rays. *Space Sci Rev* 99(1–4):85–94
- Usoskin IG, Alanko-Huotari K, Kovaltsov GA, Mursula K (2005) Heliospheric modulation of cosmic rays: monthly reconstruction for 1951–2004. *J Geophys Res-Space Phys* 110:A12108
- Webber WR, Higbie PR (2003) Production of cosmogenic Be nuclei in the Earth’s atmosphere by cosmic rays: its dependence on solar modulation and the interstellar cosmic ray spectrum. *J Geophys Res* 108(A9):1355–1365
- Webber WR, Higbie PR (2009) Galactic propagation of cosmic ray nuclei in a model with an increasing diffusion coefficient at low rigidities: a comparison of the new interstellar spectra with Voyager data in the outer heliosphere. *J Geophys Res – Part A – Space Phys* 02103:02106
- Webber WR, Lockwood JA (2001a) Voyager and Pioneer spacecraft measurements of cosmic ray intensities in the outer heliosphere: toward a new paradigm for understanding the global solar modulation process 1. Minimum solar modulation (1987 and 1997). *J Geophys Res* 106(A12):29323–29331
- Webber WR, Lockwood JA (2001b) Voyager and Pioneer spacecraft measurements of cosmic ray intensities in the outer heliosphere: Toward a new paradigm for understanding the global modulation process 2. Maximum solar modulation (1990–1991). *J Geophys Res-Space Phys* 106(12):29333–29340

Chapter 6

Instrumental Measurements of the Cosmic Radiation

6.1 Introduction

Section 5.1 outlined how Viktor Hess discovered the cosmic radiation during a balloon flight in 1912. His evidence came from two electroscopes that measured the amount of ionization produced by ionizing radiation as it passed through the gas in a closed container.

Improved versions of his instrument were soon being used – and called “ionization chambers”. At first they were unstable; however, over time they were gradually improved, and the first continuously recording ionization chambers commenced limited use in Germany and Austria in about 1928. Since then many types of cosmic ray detection systems have been designed to measure different portions of the cosmic ray spectrum, and other properties such as its composition and degree of isotropy. For the purposes of this book, we restrict ourselves to three instruments that provide us with our record of the manner in which the cosmic radiation has varied at Earth since 1933; the recording ionization chamber, the neutron monitor, and the low energy cosmic ray detectors flown on the Interplanetary Monitoring Spacecraft (IMP) and other satellites between 1968 and the present. Taken together, they provide a detailed record of the temporal changes in the cosmic ray intensity at Earth over the past >75 years. Using them, we can project both forward and back in time. Thus, we use the data from them to interpret the cosmogenic data from the period before 1933. They, together with the cosmogenic data, also allow us to estimate the radiation conditions we can expect on Earth, and in space, in the years ahead.

6.2 Ionization Chambers and Muon Telescopes

The first continuously recording ionization chambers were developed by Steinke in the late 1920s, and were installed in a number of locations in Europe and South Africa. It is not known if any of those data still exist – they could be used to extend the instrumental record further back in time if they did. One Steinke instrument made a major contribution to the debate on whether the cosmic rays were gamma rays when it was installed on a ship that sailed from Genoa to Indonesia in 1931. It provided the first high-quality measurement of the dependence of cosmic radiation intensity on geomagnetic latitude (Clay 1933). As outlined in Sects. 5.8.2 and 6.5, the existence of a “latitude effect” settled the question: cosmic rays are charged particles.

In 1936, the Carnegie Institution of Washington established continuously recording ionization chambers in the USA, New Zealand, Peru, Mexico, and later in Greenland (Compton et al. 1934; Lange and Forbush 1948). They were operated by Scott Forbush and his associates, and until 1951, they and the Steinke ionization chambers in Germany, Austria, South Africa, Sweden, and other ionization chambers in the Soviet Union and Australia provided the only record of the manner in which the cosmic radiation intensity varied with time. Forbush maintained meticulous quality control on the data; he published the bihourly data from all his instruments for over 20 years, and he published high-quality papers in important journals. As a consequence, his pioneering work is still well known, while the data from the other ground-based instruments are largely forgotten.

Using his five ionization chambers, Forbush demonstrated that the cosmic radiation at Earth is not constant, but under solar control. Sudden decreases in the intensity were observed a day or so after a large solar flare – they are now called Forbush decreases (Forbush 1938). The instruments recorded short-lived increases in the cosmic ray intensity within minutes of solar flares on 28 February and 7 March 1942, and twice later in that decade (Forbush 1946). Following the establishment of the worldwide network of neutron monitors (see Sect. 6.3), most of the ionization chambers were shut down, few being in operation after 1970. Nevertheless, they (and other balloon borne ionization chambers, see below) provide the only information for the vital period prior to the detonation of the first atomic bomb that allows calibration of the ^{14}C record to the modern era, and they also provide the only instrumental measurements of the cosmic radiation before the Sun entered its most active period from 1957 onward (Sect. 5.5). For both reasons, we outline the characteristics of these pioneering instruments.

A Carnegie Institution ionization chamber is shown in Fig. 6.2-1a, left panel (Lange and Forbush 1948). It consisted of a 19 l steel sphere filled with pure Argon at a pressure of 50 atm. The sensitive volume was surrounded by ~10 cm of lead to screen out the low-energy electrons in the nucleonic cascade (Sect. 10.2) and the response to nucleonic reactions in nearby material. The positive ions produced in the Argon by the cosmic radiation were collected onto a central electrode by a potential difference of several hundred volts with respect to the external sphere.

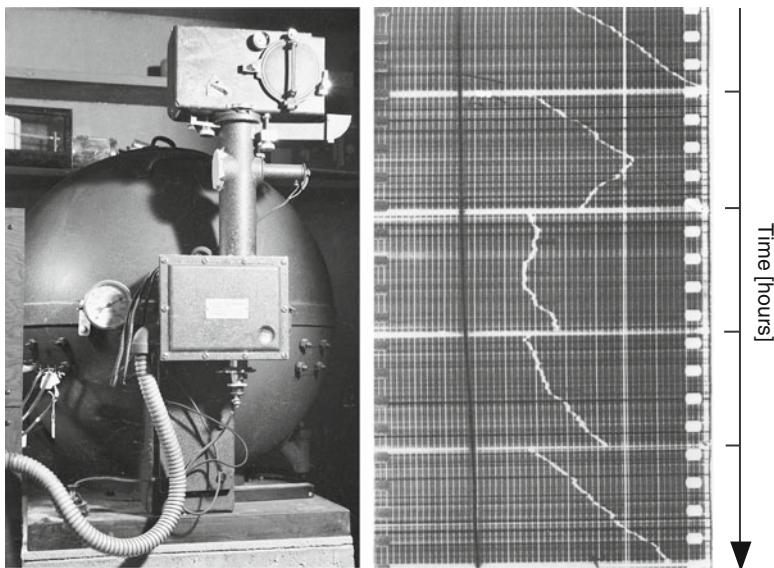


Fig. 6.2-1 *Left (a)* The ionization chamber at Huancayo, Peru, one of six operated by the Carnegie Institution of Washington from 1936 to ~1970. The principle of operation is outlined in the text. The electrometer is in the *box* in front of the chamber and the photographic recording system is immediately above. *Right (b)* Portion of the ionization chamber record from Christchurch (New Zealand) at the commencement of the Ground Level Event (GLE) on 7 March, 1942 (see Sect. 8.2.1) (Carnegie Institution of Washington)

There was a small second ionization volume enclosing part of the central electrode (the “compensation chamber”) that contained a weak radioactive source; the voltages being such that the negative ions generated in this volume were attracted to the central collection electrode. The strength of the radioactive source was adjusted so that the positive cosmic ray current to the collection electrode and the negative compensation current were in approximate balance. It was an excellent example of a “difference measurement” and resulted in the changes in the current to the collection electrode being almost totally due to variations in the cosmic ray intensity alone. In particular, the voltages, capacitances, and other instrumental characteristics were designed to ensure that the collection current was essentially invariant to changes in the voltage used, temperature, or gas pressure. The voltage on the collection electrode was measured using a very sensitive electrostatic galvanometer; a “light lever” being used to record the cosmic ray intensity on a slowly moving sheet of photographic paper (Fig. 6.2-1b, right panel). The collection electrode was earthed each hour, the rate with which the voltage drifted away from zero then yielding an accurate measurement of the cosmic ray intensity. In the event of solar cosmic ray events, time could be read to the nearest minute. At the time of writing, all the original photographic records are still in excellent condition and a portion of one is given in Fig. 6.2-1.

The portion of the data record in Fig. 6.2-1b further illustrates the operation of the instrument. Time increases downwards and cosmic ray intensity increases to the left. Starting at the top, the voltage on the central electrode was decreasing steadily as a consequence of low cosmic ray intensity as a result of a high value of the atmospheric pressure. The solar cosmic ray event commenced abruptly near the middle of the second hour, and the higher cosmic ray intensity resulted in the voltage on the central electrode increasing for the first 30 min, with the consequence that the trace on the record moved to the left. As the cosmic ray intensity returned to the pre-event level over the next 2 h, so did the trace on the record return to the drift to the right evident prior to the GLE.

There was an instrumental defect in some of the ionization chambers; however, that has had unfortunate consequences in the interpretation of ^{14}C and other data. Forbush recognized that there were long-term linear instrumental drifts from 1936 to 1956 in four of his instruments, and he de-trended the data, i.e. a linear correction was applied to eliminate the linear changes between those dates (Forbush 1954). Unfortunately, despite strong warnings in the literature that these corrections had been applied, later workers used the de-trended data to conclude that there were no long-term changes in the cosmic radiation prior to 1951. The Forbush data are widely available and they provide an invaluable record of the solar flare events, Forbush decreases, and the 11 year variations; however, it cannot be stressed too strongly – they do not tell us anything about the longer term changes in the cosmic ray intensity. Fortunately, there was another source of information that overcame this problem.

Between 1933 and 1969, Victor Neher and his colleagues flew ionization chambers to high altitudes, and concluded that there had been a long-term decrease in the cosmic ray intensity between the 1930s and 1950s (Neher 1971). Figure 6.2-2 illustrates the Neher chamber; it was small and light and was carried to high altitudes (>31 km) by several small weather balloons (Neher 1953). The later versions used radio to transmit the results back to Earth. Each ionization chamber was calibrated against a radioactive calibration source prior to flight, the instrumental accuracy over the whole 33 year interval being estimated to be $\pm 1\%$ (Neher 1971). The data from hundreds of flights led Neher to conclude that there had been a long-term decrease in the cosmic radiation during the solar cycles between 1933 and 1964. Without any other measurements to confirm this result, this conclusion was overlooked until the early 2000s when the cosmogenic ^{10}Be data were shown to exhibit a similar and quantitatively consistent decrease.

Based on that verification, the Neher data have been used to provide a long-term calibration for the Carnegie Institution data (McCracken and Beer 2007). Using modern calculations of the specific yield functions (Sect. 6.5), the ionization chamber data have been calibrated to the neutron monitor data for the period back to 1933 as shown in Fig. 6.2-3.

Finally, we make brief mention of the “muon telescopes” which were in use from the 1940s to the 1970s. During the International Geophysical Year (IGY, see below) they were widely used together with neutron monitors to study the temporal variations of the cosmic ray intensity. The IGY design consisted of three “trays” of



Fig. 6.2-2 A Neher high altitude ionization chamber. It was 15 cm in diameter and was flown to high altitudes below a cluster of small balloons. The data were sent back to Earth by radio. The pressure measuring aneroid cell can be seen to the left above the radio transmitter (California Institute of Technology)

Geiger counters, each with a sensitive area of 100×100 cm, the second and third being vertically below the top one and spaced 50 cm apart. There was a 10 cm thick layer of lead between the second and third trays to absorb low-energy muons. A high-energy cosmic ray muon passing through all three trays would produce simultaneous electrical pulses in each, which were detected by a “coincidence circuit”. The muon telescope had the advantage of a higher counting rate than the IGY neutron monitor, and it was useful at the time for studying short-term phenomena and the diurnal variation. However, the muon telescopes were relatively unstable in the long term and costly to maintain, and they were quickly superseded by the IQSY neutron monitor (see below) for the majority of cosmic ray studies.

6.3 The IGY and IQSY Neutron Monitors, and Spaceship Earth

As will be detailed in Sect. 10.2, a cosmic ray initiates a nucleonic cascade in the atmosphere that penetrates to the surface of the Earth for cosmic rays with rigidities >1 GV. In the late 1940s, John Simpson and his students at the University of Chicago developed an instrument to selectively measure the non-thermal neutron

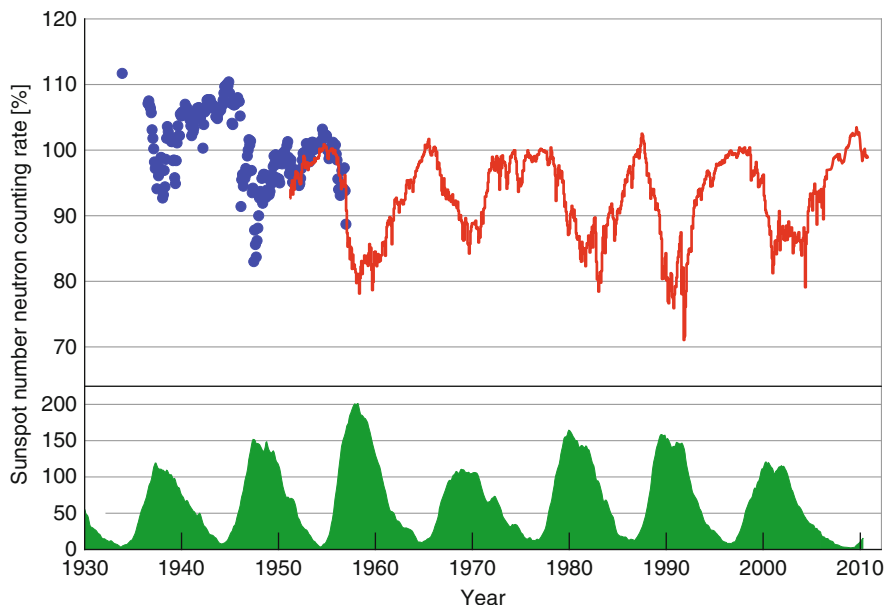


Fig. 6.2-3 The observed and estimated monthly Climax neutron counting rates (*connected line and solid dots*, respectively), and the international sunspot number, 1933–2010. Both the observed and estimated counting rates were normalized to 100% in 1954. The data 2007–2010 are from the Kiel neutron monitor

component of this nucleonic cascade (Simpson 2000). They demonstrated that it was about three to four times more sensitive to the temporal changes in the cosmic ray intensity than the ionization chambers in use at that time, and with inherently greater short- and long-term stability. They named the instrument the “neutron monitor”, and the first continuously operating instrument was installed at Climax, Colorado, USA, in 1951. There is a continuous neutron monitor record from Climax from that time (e.g. Fig. 6.2-3) to 2006, and it is one of the most commonly used long-term records of the cosmic ray intensity at Earth (<http://Ulysses.sr.unh.edu/NeutronMonitor>).

The Simpson neutron monitor was rapidly adopted as one of the two standard cosmic ray detectors recommended for use during the “International Geophysical Year” (IGY) of 1957–1958. More than 60 neutron monitors commenced operation, and this extensive network led to a rapid increase in our understanding of the manner in which the cosmic ray intensity and spectrum varies with time (Chap. 7).

Figure 6.3-1 is a photograph of an 8-counter IGY neutron monitor, while its construction is shown in Fig. 6.3-2 [from Simpson’s original paper describing the neutron monitor (Simpson et al. 1953)]. To minimize local effects, it was designed to be selectively sensitive to the high-energy neutrons in the nucleonic cascade (Fig. 10.2.1-1). A fast neutron initiates a fragmentation reaction in a lead nucleus in the large target mass. A number of evaporation neutrons are produced, and slowed



Fig. 6.3-1 An eight-counter IGY neutron monitor established in a small hut at an altitude of 725 m on the slopes of Mt. Wellington, Tasmania, in 1956. Note the large number of “vacuum tubes” that were needed to perform the simplest of tasks prior to the advent of solid-state electronics and the digital computer. The student checking the instrument is one of the authors of this book (collection McCracken)

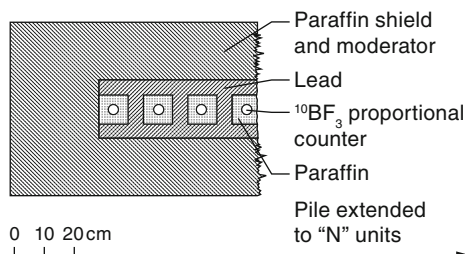


Fig. 6.3-2 A cross-section view of the IGY neutron monitor. In practice, monitors would contain from 4 to 12 counters, depending on the generosity of the funding agency. This instrument was designed to respond selectively to the fast neutrons in the nucleonic cascade as outlined in the text. A 12-counter neutron monitor (such as operated at Climax, Colorado) used 3,000 kg of lead, and an external shield of 1,400 kg of paraffin wax

down (“moderated”) by the paraffin wax surrounding the $^{10}\text{BF}_3$ proportional counters. A ^{10}B nucleus in the boron trifluoride counter gas captures a slow neutron, disintegrates (in nuclear physics shorthand, $^{10}\text{B}(n,\alpha)^7\text{Li}$), and the resulting alpha particle and ^7Li nuclei produce a charge pulse in the counter that is much larger than those due to the other components of the nucleonic and electromagnetic cascade initiated by the original cosmic ray. To reduce the sensitivity of the monitor to fast neutrons produced in (movable) matter, snow, etc. outside the monitor, the target mass is surrounded by a thick shield of paraffin wax.

The flux of neutrons in the nucleonic cascade increases rapidly with altitude, increasing by a factor of ~ 10 between sea level and 3,000 m. The statistical

fluctuations in the neutron-counting rate are essentially Gaussian; therefore the percentage standard deviation in the hourly data is given by $100/N^{0.5}$, where N is the hourly counting rate. The percentage standard deviation therefore decreases by a factor of three if a neutron monitor is moved from sea level to an altitude of 3,000 m, say. The counting rate of the largest IGY neutron monitor (12 neutron counters) was relatively low at sea level ($\sim 50,000$ counts/h at high latitudes) yielding a standard deviation of 0.5%, and this was barely adequate for the investigations of interest during the IGY. Further, neutron monitors were relatively expensive for those times, and consequently many 4- and 6-counter neutron monitors were constructed. For both the reasons, a substantial number of the IGY neutron monitors were established at high altitudes (e.g. Climax).

In the early 1960s, Hugh Carmichael, at the Deep River laboratories of the Atomic Energy of Canada, developed a much larger neutron monitor that became the standard neutron monitor for the “International Quiet Sun Year” (IQSY) of 1964 (Carmichael et al. 1968). The target mass was increased by a factor of ~ 10 ; the counters increased in length and diameter by a factor of ~ 2.5 ; the paraffin wax was replaced by more stable polyethylene; and the electronics were based on semiconductor devices, yielding greater stability and reliability than the “vacuum tube” electronics used in most IGY monitors. To a large degree, the IQSY neutron monitor replaced the IGY monitors of the 1950s. The “standard” IQSY monitor used 18 of the large proportional counters, initially operated as three totally independent units to permit the detection and correction for sensitivity changes. Until ~ 1990 , Carmichael and his colleagues operated a 48 counter IQSY monitor at Deep River, Canada, and this remains one of the most sensitive records of the time dependence of the cosmic ray intensity in existence. The cosmic ray literature uses the code 18NM64, 9NM64, etc., to describe a neutron monitor, where the first number indicates the number of counters in use in Carmichael’s design of 1964.

The IQSY neutron monitor (also called the “super” neutron monitor) has continued to evolve with time. In some, the $^{10}\text{BF}_3$ proportional counters have been replaced with proportional counters using ^3He as the filling gas. In these, the slow neutrons interact with a ^3He nucleus in the reaction $^3\text{He}(n,p)^3\text{H}$ to yield an electrical pulse similar to that obtained in a $^{10}\text{BF}_3$ counter. In some, all the counters are operated independently to improve the detection and correction of instrumental changes. None of these changes have influenced the sensitivity of the neutron monitors to a significant degree (Clem and Dorman 2000).

For primary galactic cosmic rays with rigidity < 1 GV (proton energy ≤ 0.414 GeV) the nucleonic cascade becomes too small to be detected with any efficiency by a neutron monitor at sea level. As a consequence, the counting rate is the same for all vertical cut-off rigidities < 1 GV. This can be seen in Fig. 6.3-3, which displays the “latitude curve” obtained as a neutron monitor is moved from the equator to the polar regions. The abrupt change in slope at 1 GV is called the “knee” of the latitude curve. For geomagnetic latitudes $< 60^\circ$ (vertical cut off > 1 GV) the counting rate steadily decreases as the cut-off rigidity increases to ~ 17 GV near the equator (Sect. 5.8.2).

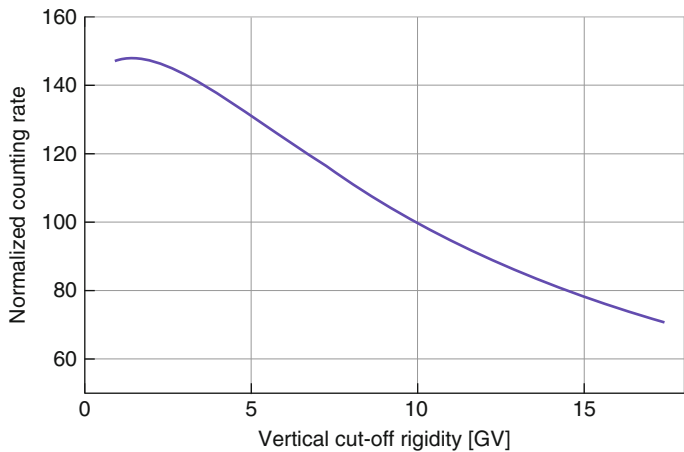


Fig. 6.3-3 The “latitude effect” of a neutron monitor at sea level. These data were obtained by the South African cosmic ray group on board several ships that sailed to Antarctica, Japan, and the USA (Moraal et al. 1989), and are plotted against vertical cut-off rigidity (Sect. 5.8.2)

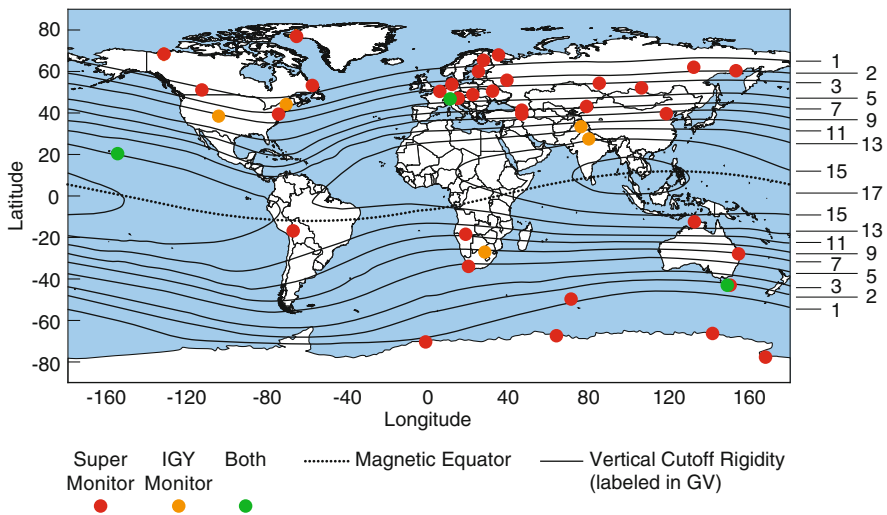


Fig. 6.3-4 Worldwide distribution of neutron monitors in 1999. The *colour* indicates the monitor type. The *contours* show the vertical cut-off rigidity. A number of these instruments have ceased operating up to the time of writing [After (Simpson 2000)]

Figure 6.3-4 displays the worldwide distribution of neutron monitors, superimposed upon a map and the contours of vertical cut-off rigidity.

A neutron monitor located within about 25° of the magnetic poles has a unique property in that it only responds to cosmic rays coming from a very limited set of direction in space (McCracken 1962). To exploit this property to provide a detailed understanding of the directional characteristics of the cosmic radiation, a network

of IQSY monitors called “Spaceship Earth” was established by the University of Delaware and several international collaborators in 2000 (Moraal et al. 2000). The locations were carefully chosen to provide an even distribution of the viewing directions over the celestial sphere, with particular emphasis on the study of the directional characteristics of the cosmic rays occasionally generated by the Sun (GLE, Sect. 8.2.1) and the Forbush decrease (Sect. 7.4) (<http://neutronm.bartol.udel.edu>).

The sensitivities of the IGY and IQSY neutron monitors to the temporal and spectral changes in the cosmic radiation are essentially the same, and their data can be used interchangeably for most purposes (Clem and Dorman 2000). There are small differences between the specific yield functions of sea level, and high altitude neutron monitors, and for some investigations, it is necessary to recognize these differences (see Sect. 6.6).

6.4 Satellite Borne Detectors

As shown by the equation (6.5-1), the ionization chamber and the neutron monitor both integrate over the whole cosmic ray spectrum, and over all nuclear species (protons, helium nuclei, etc.). In other words, a single instrument cannot provide any information regarding changes in the cosmic ray spectrum. Changes in the energy (or rigidity) spectrum during the 11-year variation, say, were determined by comparing the amplitudes observed by instruments with different cut-off rigidities (Sect. 5.8.2). Neither instrument could ever provide any information about the behaviour of the different nuclear species in the cosmic radiation incident on Earth.

This situation changed completely with the advent of the space age. Instruments were quickly developed that could measure both the energy, and the charge, of each cosmic ray that entered the detector. It is these instruments that have provided the detailed spectral information that is now used to interpret cosmogenic data from the past (the Rosetta stone principle), and to compute accurate ionization rates and radiation doses in space weather applications.

Many of these charge and energy sensitive detectors were (and still are) similar to that shown in Fig. 6.4-1a (McGuire et al. 1986). A cosmic ray is detected when it first goes through the thin detector “D” (the dE/dx detector) and then comes to the end of its range (i.e. it stops) in the thick (total E) detector, “E”. Using detectors whose outputs are proportional to the amount of energy deposited (first scintillators, later solid state detectors), the total energy of the particle is determined by summing the energy deposited in the two detectors. Figure 6.4-1b displays the result of plotting dE/dx versus E for each cosmic ray that stops in the detector. The energy loss in the thin detector is proportional to the square of the charge of the cosmic ray divided by its energy (see Eq. 10.2.2-1), and consequently the various nuclei define a series of bands, as is clearly seen in the figure. This allows the nuclear species of each cosmic ray to be identified. Over time, these detectors have been refined to the

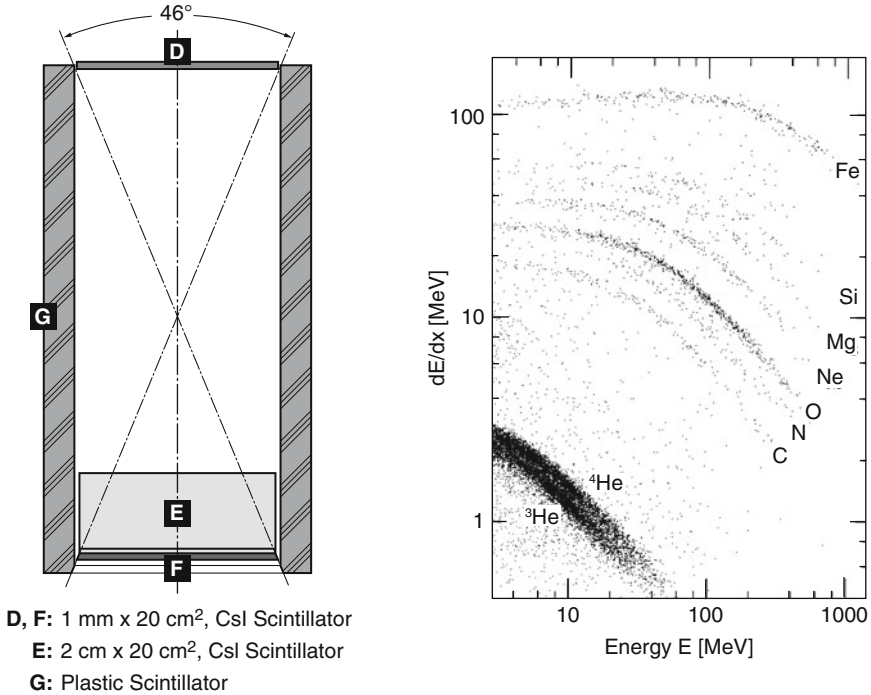


Fig. 6.4-1 (a) Schematic diagram of the $(dE/dx, E)$ cosmic ray detectors flown on the Interplanetary Monitoring Platform (IMP) satellites (see Fig. 4.1) and which provided a continuous record of the low energy cosmic radiation intensity between 1962 and 2006. (b) Illustrating the manner in which the different cosmic ray nuclei define separate bands on the $(dE/dx, E)$ plane, allowing the energy spectrum of each component to be measured (McGuire et al. 1986; Reames et al. 1997). These data are from the low-energy version of the instrument shown in panel (a)

point that they can easily resolve the different isotopes of the various components of the cosmic radiation (e.g., ^3He from ^4He ; and between ^7Be , ^9Be , ^{10}Be). Cosmic rays that trigger either of the guard detectors “F” and “G” in the diagram are rejected since their inclusion would degrade the energy and charge resolution in Fig. 6.4-1b.

Starting in the 1960s, Frank McDonald and his co-workers at the Goddard Space Flight Centre of NASA, and John Simpson at the University of Chicago flew this type of detector on a number of the Interplanetary Monitoring Platform (IMP) and other satellites, and until 2006 these provided an essentially continuous measurement of the cosmic ray proton, helium, and heavier nuclei in six narrow energy bands between 130 and 450 MeV/nucleon. Similar instruments were flown on spacecraft (Pioneers 10 and 11, and Voyagers 1 and 2) that travelled to and beyond the termination shock; and others in polar orbits around the Sun. Together, all these data have allowed us to understand the cosmic ray modulation processes, and thereby provide the “Rosetta Stone” we use to interpret the cosmogenic data. Following the demise of the IMP8 spacecraft in 2006, measurements made by

another detector on the Advanced Composition Explorer (ACE) spacecraft have been calibrated to allow the IMP data series to be continued into the future.

In recent years, even more accurate and sensitive detectors have been flown on satellites. Time-of-flight versus E detectors have been flown on WIND, ACE, and STEREO. Cherenkov detectors have extended the energy range to ~ 700 MeV on a number of satellites (e.g. the HEPAD instrument on GOES). Perhaps of greatest potential are the large, multi-detector magnetic spectrometers (PAMELA, AMS (Alpha Magnetic Spectrometer)) that provide greatly improved measurements of the cosmic ray spectrum to >100 GeV. We anticipate that these will lead to better estimates of the LIS, and to the ultimate calibration of the modulation function (Sect. 5.7.4). Finally, we mention the suite of instruments on GOES that provides a continuous record of solar energetic particle (SEP) events- see <http://spidr.ngdc.noaa.gov/spidr/>

6.5 Latitude Effects and the Yield Functions

In the late 1920s, it was recognized that charged cosmic rays would be affected by the geomagnetic field and the concept of the Störmer cut-off (Sect. 5.8.2) was enunciated. At that time controversy still raged whether the cosmic rays were charged particles or gamma rays. In attempts to answer this question, several ionization chambers were installed on ships (Sect. 6.2), and these demonstrated that there was a $\sim 10\%$ dependence of ionization current on latitude (i.e. upon geomagnetic cut-off rigidity).

From the beginning of Simpson's investigations of the nucleonic component in 1948 using high-flying aircraft, it was clear that it exhibited a much larger latitude effect than the ionization chamber (Fig. 6.3-3). Simpson and his students introduced the concept of the "specific yield function", $S(E)$ or $S(P)$, to quantify the sensitivity of the instruments to the cosmic radiation as a function of energy or rigidity (Simpson et al. 1953). Thus for any type of detection system they wrote the equivalent of

$$X(E_c, t) = \sum \int_{E_c}^{\infty} J(E, t) S(E) dE \quad (6.5-1)$$

where

$X(E_c, t)$ is the cosmic ray measurement at a location where the proton cut-off energy is E_c , at time t ,

$J(E, t)$ is the differential spectrum of the cosmic radiation near Earth for the proton, He, and heavy components of the cosmic radiation,

$S(E)$ is the Specific Yield function, and integration is from the geomagnetic cut-off energy E_c to infinity, and the summation is over the proton, alpha and heavies in

the cosmic ray spectrum. Note that the cut-off energies for the He and heavy components are approximately half those of the proton (Sect. 5.8.2, and Eq. 5.3.7b).

They and others used the observed latitude effects to estimate $S(E)$, and this became a vital concept in the understanding of the data observed by the worldwide network of neutron monitors and muon detectors during the International Geophysical Year, 1957–1958. This led to the first quantitative measurements of the rigidity dependence of the 11-year and the other modulation mechanisms outlined above.

There were no experimental data that would allow the specific yield functions to be estimated above 17 GeV (the highest geomagnetic cut-off energy at the equator, see Fig. 5.8.2-4). Various estimates were used to overcome this problem; however, they introduced unquantifiable uncertainties into the investigations. This problem was not solved until the 1990s, when mathematical models such as GEANT (see Sect. 10.2) were developed that used known nuclear cross-sections to compute the details of the interaction and propagation of the cosmic ray cascade through the atmosphere. Those programs have subsequently allowed us to compute the specific yield functions for all types of cosmic ray instruments (or measurements) for all energies (Clem and Dorman 2000), as well as computing the production rates of the cosmogenic radionuclides (Sect. 10.3).

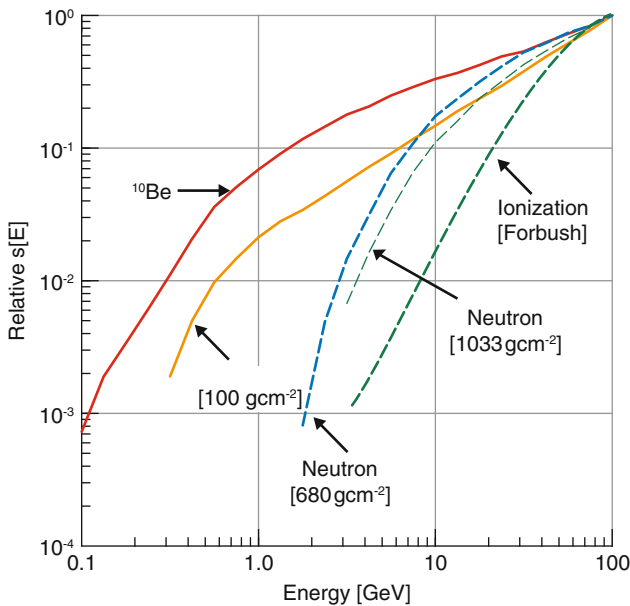


Fig. 6.5-1 The specific yield functions for cosmic ray protons applicable to five different cosmic ray measurements. The upper neutron monitor curve applies to the Climax neutron monitor; the lower to an instrument at sea level. Note that the neutron monitors rapidly decrease in sensitivity below ~ 3 GeV, while ^{10}Be is sensitive to an order of magnitude lower proton energies. The ^{14}C curve is essentially similar to that of ^{10}Be . The specific yield functions for He and heavier cosmic rays are similar when expressed as functions of energy per nucleon. Note that these curves, unlike the next Fig. 6.5-2, apply to both the galactic and solar cosmic radiation

Fig. 6.5-2 The response functions for galactic cosmic rays for the 130–450 MeV IMP measurements; ^{10}Be ; high latitude sea level neutron monitors, and ionization chambers. The ^{14}C curve is similar to that of ^{10}Be . These curves include the contributions made by the proton, He, and heavy components of the cosmic radiation. As discussed in the text, this figure does not apply to the cosmic radiation produced by the Sun (see Sect. 8.3-1)

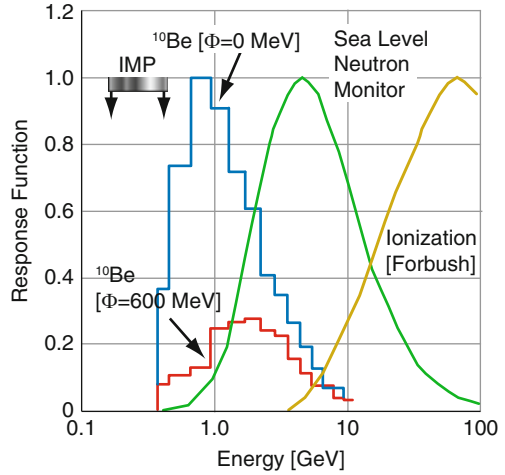


Figure 6.5-1 displays modern values of the specific yield functions, and provides an extremely clear demonstration of the differences in energy sensitivity between the various types of measurements (McCracken and Beer 2007). We will discuss these below.

Equation (6.5-1) is frequently rewritten thus

$$X(E_c, t) = \sum \int_{E_c}^{\infty} R(E, t) dE \quad (6.5-2)$$

where $R(E, t) = J(E, t) * S(E)$ is called the “Response Function” and is plotted in Fig. 6.5-2. This shows the sum of the contributions made to the observations by protons, He, and heavy cosmic rays, as a function of cosmic ray energy per nucleon. Thus while all of the sensitivities (the specific yield functions) are dropping quickly towards lower energies, we have seen in Sect. 5.4 that the galactic cosmic ray spectrum near Earth varies roughly as $E^{-2.37}$ above about 3 GeV. For the galactic cosmic radiation, these two effects largely compensate and the response functions in Fig. 6.5-2 show the result. We stress, however, that the response function for solar cosmic rays is quite different, extending to much lower energies (a consequence of a much steeper differential spectrum, see Sect. 8.3.1). In practice, the specific yield function is the more useful for calculations, while the response function provides a very clear visual understanding of the energies that contribute to the several measurements, as we now summarize.

1. Figures 6.5-1 and 6.5-2 show that the ground level ionization chamber is insensitive to primary galactic cosmic rays with $E < 3$ GeV/nucleon, and has its peak response at ~ 70 GeV/nucleon. Figure 6.5-1 shows that it is systematically less sensitive than all other measurements at higher energies. The cosmic modulation varies roughly as E^{-1} , and this curve therefore explains why the

ionization chamber observes smaller variations than all the other techniques summarized here.

2. Figure 6.5-1 shows that the neutron monitor is intermediate in sensitivity to the ionization chamber and the cosmogenic measurements. Figure 6.5-2 shows that it is essentially insensitive to galactic cosmic rays with $E < 1$ GeV/nucleon, and has its peak response at ~ 4.5 GeV/nucleon. Figure 6.5-1 shows that a neutron monitor at a high altitude (e.g. Climax) is somewhat more sensitive to the lower energies than one at sea level, and as a consequence the amplitudes of the various cosmic ray variations are weakly dependent upon altitude.
3. For historical purposes, Fig. 6.5-1 shows that the Neher ionization chamber was sensitive to quite low energies (provided the geomagnetic cut-off was low) (McCracken and Beer 2007). Since they were flown at high altitudes, quite low energy cosmic rays could reach the chambers before they came to the end of their range. The graph given is for an atmospheric depth of 100 g cm^{-2} . The $S(E)$ curve is higher and extends to even lower energies for higher altitudes.
4. Figure 6.5-1 shows that the sensitivity of cosmogenic ^{10}Be and ^{14}C extends to low energies (< 100 MeV/nucleon) while the response functions (Fig. 6.5-2) peak between 0.9 and 1.5 GeV/nucleon, depending on the value of the modulation function, Φ . The figures make it clear that the cosmogenic data are significantly more sensitive to the lower energies than the neutron monitor, and that the temporal variations will be larger. The next section will quantify that difference.
5. Finally, as shown in Fig. 6.5-2, the energy sensitivity of the IMP detector is sharply defined by the electronic thresholds employed; in the case shown they were 130–450 MeV/nucleon. The inverse dependence of the modulation upon energy means that this instrument sees much larger variations than either the cosmogenic data or the neutron monitor.

6.6 Inter-calibration of the Different Cosmic Ray Records

By the late 1980s, it was clearly evident that the amplitudes of the cosmic ray variations in different data sets were strikingly different. Thus for the 11 year variation (see Sect. 7.2), they were approximately: Forbush ionization chamber = 5%; neutron monitor = 20%, ^{10}Be = 40% and IMP 130–450 MeV/nucleon = 70%. As discussed in Sect. 5.7, it is well known that the cosmic ray modulation is a relatively strong function of particle energy (varying approximately as E^{-1}). This, together with the different energy sensitivities discussed in the previous section, explain these differences in general terms. We now describe a method that allows us to make quantitative conversions between these various data types. This is important, since there is no single record that extends from 10,000 years ago, up to the present day.

Using (6.5-1) and the specific yield functions such as shown in Fig. 6.5-1, the outputs of the various detectors (and the cosmogenic production) were computed for a series of values of the modulation function, Φ (see Sect. 5.7.3). In most cases, the ratio $X(\Phi)/X_{\text{LIS}}$ was computed where $X(\Phi)$ was the value computed using (6.5-1)

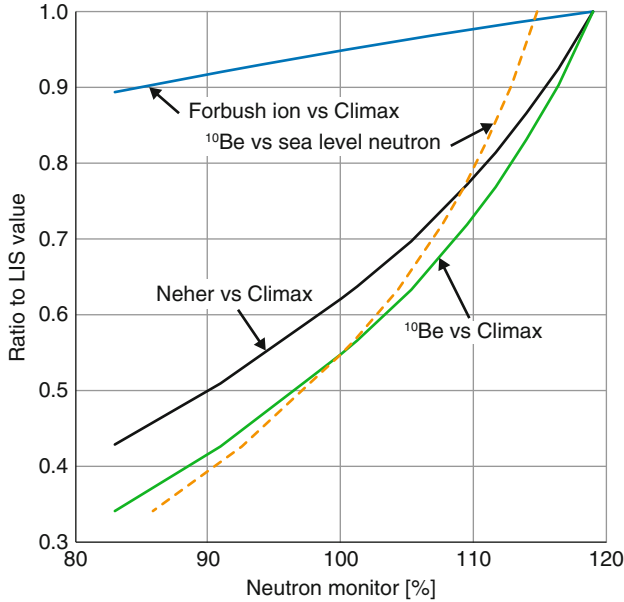


Fig. 6.6-1 The inter-calibration curves between a number of different types of cosmic ray data. The ^{14}C curves are similar to those for ^{10}Be

for a modulation function of Φ , and X_{LIS} was the value for $\Phi = 0$ (i.e., corresponding to the local interstellar spectrum). Figure 6.6-1 displays the instrumental inter-calibration curves obtained by plotting the computed outputs for the various data types against the estimated output of the Climax neutron monitor (but at the same cut-off rigidity as at present, see “pseudo-Climax” below) (McCracken and Beer 2007). By long established tradition, the Climax neutron monitor counting rate is taken to be 100% for August 1954, that being the maximum intensity attained during the sunspot minimum of 1954. The curves show that the computed neutron monitor response for $\Phi = 0$ is 119.5% for the pseudo-Climax record, and 115% for a high latitude neutron monitor at sea level. The high altitude response is higher because of the higher sensitivity evident in Fig. 6.5-1.

Figure 6.6-1 shows that with the exception of Forbush ion versus neutron monitor, the inter-calibration curves are quite non-linear. For example, the gradient of the ^{10}Be and pseudo-Climax neutron curve changes by a factor of 2.7 between the periods near the Maunder and Dalton Minima ($0 < \Phi < 200$ MeV) and the period since 1951 ($500 < \Phi < 1,000$ MeV). This shows that the use of a linear regression derived from the period after 1951 to estimate the neutron monitor response for no modulation ($\Phi = 0$ MeV) would yield an estimate of 136%, compared to 119% for the computed non-linear relationship. All the inter-relationships between ^{10}Be and ^{14}C , the neutron monitors as a class, the high-altitude ionization chambers, and the satellite data exhibit similar or greater errors if

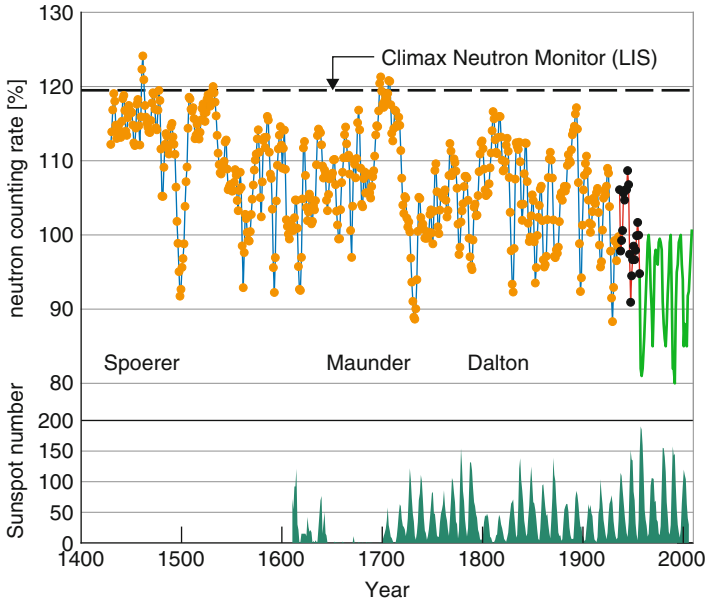


Fig. 6.6-2 The estimated counting rate for a hypothetical (“pseudo-Climax”) neutron monitor at the cut-off rigidity of Climax, Colorado, from 1428–2010. The *yellow dots* and *blue line* are derived from the ^{10}Be data; the *black dots* from the Forbush ionization chamber data; and the *green connected lines* are from the Climax and Hermanus (after 2006) neutron monitors. The *horizontal dashed line* is the counting rate that would be observed if the local interstellar spectrum of galactic cosmic rays were incident on Earth. The intervals labelled Spoerer, Maunder and Dalton were “Grand Minima” of solar activity (Fig. 5.5-2)

a linear regression is used. The non-linear inter-relationships in Fig. 6.6-1 are used throughout the rest of this book.

Figure 6.2-2 is the resulting inter-calibrated cosmic ray record for the interval 1428–2005. It used the Neher ionization chamber neutron monitor inter-calibration curve in Fig. 6.6-1 for the interval 1933–1951. Anticipating our later discussions about the ^{10}Be observations, it used the ^{10}Be –neutron monitor inter-calibration curve for 1428–1933. It is important to note that normalization or “curve matching” was not used to obtain the continuity between the ^{10}Be , Forbush and neutron data in the figure. The plotted values were obtained solely using the inter-calibration curves in Fig. 6.6-1; that is, they are a direct outcome of the GEANT computations outlined above, and the understanding of the cosmic ray modulation process obtained over the past 50 years.

As outlined in Sect. 5.8.2, the long-term changes in the geomagnetic field mean that the geomagnetic cut-offs vary with time. For example, the decreasing strength of the geomagnetic dipole and polar wander mean that the geomagnetic cut-off of Climax was ~ 6 GV in 1600 AD, compared to 3.15 GV in 1990. It is still changing quite rapidly. To save complication, the curves in Fig. 6.6-2 were computed for the

Climax neutron monitor held at a constant geomagnetic cut-off of 3.15 GV. We refer to these as the “pseudo-Climax” data.

Clearly, it would be very desirable to obtain further validation of the inter-calibration curves by making direct comparisons between the ^{14}C and ^{10}Be measurements and the neutron monitor record since 1951. The atomic tests starting in the 1940s injected large quantities of ^{14}C into the atmosphere, making direct comparisons meaningless in the case of ^{14}C . Prior to that, the inter-calibration is complicated by the increasing concentration of anthropogenic CO_2 in the atmosphere (devoid of ^{14}C .) For ^{10}Be , it has been difficult to obtain good sample recovery from boreholes drilled into lightly consolidated firn (the past 30 years, say) and sampling has often ceased about 1960–1970 for this reason. In very long cores (e.g. the GRIP, GISP, and EPICA cores extending $>100,000$ years into the past), technical considerations dictate that sampling commenced ~ 300 – $1,000$ years ago. However, recent programs where samples were taken from the sides of pits or loading ramps bulldozed into the ice are providing encouraging results. A pilot study undertaken in Queen Maud Land in the Antarctic that sampled the ice for the interval 1994–2003 (Moraal et al. 2005) yielded a preliminary value of 2.31 ± 0.08 for the (^{10}Be /neutron monitor) ratio, compared to a prediction of 2.57 based on Fig. 6.6-1. Repeat measurements for that period and a longer time span are desirable to complete the inter-comparison. Nevertheless, the results of the pilot study provide confidence that the inter-calibration curves are accurate to within 10% for modulation functions (Φ) between 450 and 1,200 MeV. The inter-calibration curves are based on the nuclear cross-sections used in the GEANT simulations without reference to the cosmic ray data; so this agreement for $450 < \Phi < 1,200$ MeV provides confidence that the inter-calibrations will have a similar accuracy for $0 < \Phi < 450$ MeV.

6.7 Cosmic Ray Archives

Finally we summarize some of the more important cosmic ray archives to assist further research.

1. *The ionization chamber records of the Carnegie Institution of Washington* (The Forbush records). The data from Cheltenham (USA), Christchurch (New Zealand), Huancayo (Peru), and Godhavn (Greenland) were published covering the years 1936–1968. They consist of bihourly and daily average data. As noted in Sect. 6.2, they had been detrended to remove long-term drifts from the data. They were published as four reports:

I. Lange and S.E. Forbush, Cosmic ray results from Huancayo observatory, Peru, June 1936–December, 1946, and other observatories. Res. of the Department of Terrestrial Magnetism, Vol XIV, Carnegie Institution of Washington, Washington, DC, 1948.

I. Lange and S.E. Forbush, Cosmic ray results from Huancayo observatory, Peru, January 1946–December, 1955, and other observatories. Res. of the Department of

Terrestrial Magnetism, Vol 20, Carnegie Institution of Washington, Washington, DC, 1957.

L. Beach and S.E. Forbush, Cosmic ray results from Huancayo observatory, Peru, January 1956-December, 1959, and other observatories. Res. of the Department of Terrestrial Magnetism, Vol 21, Carnegie Institution of Washington, Washington, DC, 1961.

I. Lange and S.E. Forbush, Cosmic ray results from Huancayo observatory, Peru, January 1960 -December, 1968, and other observatories. Res. of the Department of Terrestrial Magnetism, Vol 22, Carnegie Institution of Washington, Washington, DC, 1969.

2. *The Neher ionization chambers*. These data were never published as a single archive. On the other hand, the detailed ionization rates versus atmospheric depth for many flights between 1936 and 1970 were tabulated in the many papers Neher published over this period. The results from some ten flights at roughly annual intervals during the sunspot cycle 1954–1965 are given in three papers (Neher 1967, 1971) (Neher et al. (1953), and the references therein, provide the results from a large number of flights from locations ranging from Texas to the Arctic between 1936 and 1941).

References

- Carmichael H, Bercovitch M, Shea MA, Magidin M, Peterson RW (1968) Attenuation of neutron monitor radiation in the atmosphere. *Can J Phys* 46(10P4):1006–1013
- Clay J (1933) The corpuscular nature of the cosmic radiation and the geomagnetic effect. *Naturwissenschaften* 21:43–44
- Clem JM, Dorman LI (2000) Neutron monitor response functions. *Space Sci Rev* 93(1–2):335–359
- Compton AH, Wollan EO, Bennett RD (1934) A precision recording cosmic-ray meter. *Rev Sci Instrum* 5(12):415–422
- Forbush SE (1938) On world-wide changes in cosmic-ray intensity. *Phys Rev* 54(12):975–988
- Forbush SE (1946) Three unusual cosmic-ray increases possibly due to charged particles from the sun. *Phys Rev* 70(9–10):771–772
- Forbush SE (1954) World-wide cosmic-ray variations, 1937–1952. *J Geophys Res* 59(4):525–542
- Lange I, Forbush SE (1948) Cosmic-Ray Results From Huancayo Observatory, Peru, June 1936 – December, 1946. In: Res. of the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, D. C
- McCracken KG (1962) Cosmic-ray flare effect. 1. Some new methods of analysis. *J Geophys Res* 67(2):423–434
- McCracken KG, Beer J (2007) Long-term changes in the cosmic ray intensity at Earth 1428–2005. *J Geophys Res-Space Phys* 112:A10101
- Mcguire RE, Vonrosenvinge TT, McDonald FB (1986) The composition of solar energetic particles. *Astrophysical Journal* 301(2):938–961
- Moraal H, Potgieter MS, Stoker PH, Vanderwalt AJ (1989) Neutron monitor latitude survey of cosmic-ray intensity during the 1986/1987 solar minimum. *J Geophys Res-Space Phys* 94(A2):1459–1464
- Moraal H, Belov A, Clem JM (2000) Design and co-ordination of multi-station international neutron monitor networks. *Space Sci Rev* 93(1–2):285–303

- Moraal H, Muscheler R, du Plessis L, Kubik PW, Beer J, McCracken KG, McDonald FB (2005) ^{10}Be concentration in the ice shelf of Queen Maud Land, Antarctica. *S Afr J Sci* 101(5–6):299–301
- Neher HV (1953) An automatic ionization chamber. *Rev Sci Instrum* 24(2):99–102
- Neher HV (1967) Cosmic-ray particles that changed from 1954 to 1968 to 1966. *Journal of Geophysical Research* 72:1521–1539
- Neher HV (1971) Cosmic rays at high latitudes and altitudes covering 4 solar maxima. *J Geophys Res* 76(7):1637–1651
- Neher HV, Peterson VZ, Stern EA (1953) Fluctuations and latitude effect of cosmic rays at high altitudes and latitudes. *Phys Rev* 90:655–672
- Reames DV, Barbier LM, VonRosenvinge TT, Mason GM, Mazur JE, Dwyer JR (1997) Energy spectra of ions accelerated in impulsive and gradual solar events. *Astrophys J* 483(1):515–522
- Simpson JA (2000) The cosmic ray nucleonic component: the invention and scientific uses of the neutron monitor (Keynote lecture). *Space Sci Rev* 93(1–2):11–32
- Simpson JA, Fonger W, Treiman SB (1953) Cosmic radiation intensity-time variations and their origin. 1. Neutron intensity variation method and meteorological factors. *Phys Rev* 90(5):934–950

Chapter 7

Time Variations of the Cosmic Radiation

7.1 Introduction and Atmospheric Effects

From the 1920s, there was a great interest in exploring the source of the cosmic rays. Were they coming from the Sun, or from our galaxy? To determine this, investigators sought to determine if the intensity depended on solar or sidereal (astronomical) time. A number of recording ionization chambers were established in the late 1920s to determine this. It was soon realized that the intensity depended strongly on the atmospheric pressure (i.e. the amount of absorber above the ionization chamber) and also on the temperature throughout the atmosphere.

By the early 1930s, it was clear that there was a small diurnal (24-h period) variation in the ionization chamber data; however, it was not yet clear whether that was of atmospheric origin. In 1937 Scott Forbush observed simultaneous, large decreases in the intensities observed by his ionization chambers in the USA, Mexico City, Peru, and New Zealand soon after the commencement of a very large magnetic storm (Forbush 1938). Viktor Hess, the discoverer of cosmic rays, confirmed these with observations made with a Steinke ionization chamber on the Hafelekar, a high mountain in Austria. This was the first clear proof that the intensity of the cosmic radiation reaching the top of the atmosphere was not constant. Soon after, Forbush demonstrated that part of the small 0.5% diurnal periodicity was of extraterrestrial origin. Then in 1942 he observed two large bursts of cosmic rays from the Sun (Forbush 1946). Then in 1954 he showed that the cosmic ray intensity exhibited an 11-year periodicity that was out of phase with the solar activity cycle (Forbush 1954). Almost single-handedly, Forbush had discovered three of the most important time variations in the cosmic radiation, and had clarified the origin of the fourth. Several of these variations have a significant influence upon the production rate of the cosmogenic radionuclides, and the intensity of the cosmic radiation throughout the solar system, and we describe their properties in the following sections. For completeness we briefly summarize some of the other smaller variations. The production of cosmic rays by the Sun is covered in Chap. 8.

As noted above, the diurnal variation was soon recognized to be partially due to diurnal changes in the atmospheric pressure and in the temperature throughout the atmosphere. It soon became obvious that the latter effect was largely a consequence of the short mean lifetime of the muon ($2.2 \mu\text{s}$); in a warm atmosphere the muons observed by the ionization chambers were produced higher above the Earth than if it were cold, and more of them decayed before they reached Earth. The daily variation of atmospheric temperature therefore resulted in a higher ionization chamber readings at night compared with daytime. For the same reason it was also higher in winter than in summer.

Neutron monitor data are also quite strongly affected by atmospheric pressure, but not by the temperature throughout the atmosphere. This is no surprise, as we can see in Sect. 10.2 the nucleonic cascade has a mean free path of about 160 g cm^{-2} in the atmosphere and the intensity therefore exhibits an exponential dependence on pressure that can amount to $\pm 25\%$ in some locations. Section 10.2 also shows that the attenuation of the nucleonic cascade only depends on the amount of matter traversed (in g cm^{-2}), and not on its distribution above the earth's surface, so atmospheric temperature has negligible influence on the neutron monitor counting rate. Most of the neutron monitor data in the literature and on the Internet have had the pressure effects removed by an exponential correction for the observed pressure by the operators of the instruments. The details of these corrections are beyond the scope of this book; suffice to say that in practice the correction mean free path (λ) depends on latitude (longer near the equator), altitude, and on the shape of the cosmic ray spectrum. For example, λ may be as low as 100 g cm^{-2} for solar cosmic rays (Sect. 8.3.1), and it varies slightly during the solar cycle as the lower energy particles are removed selectively by the solar modulation. Thus λ is slightly larger at sunspot maximum than at minimum.

The atmospheric effects upon ionization chambers and neutron monitors are quite local, being restricted to the atmosphere within about 45° of the vertical above the instrument. The situation for the cosmogenic radionuclides such as ^{10}Be is totally different; here the data we obtain from the polar caps, for example, are influenced by the worldwide circulation of the atmosphere as discussed in detail in Chap. 13. Finally, the atmosphere has no effect upon ^{14}C observed in biological materials; however, those data are seriously influenced by other factors as discussed in Chap. 10.

7.2 The Eleven- and Twenty-Two-Year Variations

Figure 6.2-3 shows the neutron data observed at Climax, Colorado, from 1951 to the present, together with the estimated neutron-counting rate based on the data from the Carnegie Institution, and the Neher ionization chambers, 1933–1956 (Sects. 6.5 and 6.6). Figure 7.2-1 shows data from two neutron monitors, high-altitude balloon measurements, IMP data, and sunspot numbers for four solar cycles 1954–1996. Figure 7.2-2 displays the proton and helium fluxes observed by several

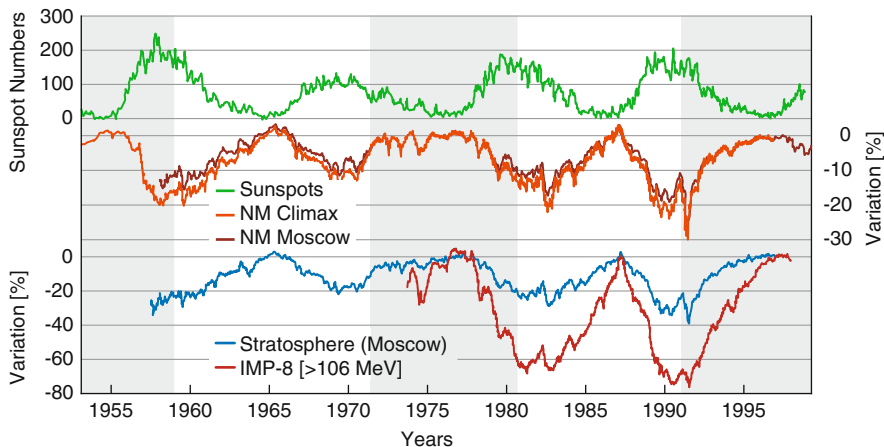


Fig. 7.2-1 The 11- and 22-year variations in the cosmic radiation (Belov 2000). *Top panel:* sunspot numbers. *Second panel:* two neutron monitors. *Third panel:* stratospheric and satellite data. Note that the cosmic ray intensity is high when solar activity is low (i.e. low sunspot numbers), and vice versa. Note also that the amplitude of the cosmic ray effect varies between $\sim 20\%$ for the neutron monitors, $\sim 30\%$ for the stratospheric measurements, and 70% for the IMP data. The grey shading indicates that the solar magnetic dipole was positive polarity; the intervening (unshaded areas) negative polarity (see explanation in text, and also in Sect. 5.7.5)

spacecraft between 1971 and 2009 (McDonald 2000). It will be noted that there are persistent large amplitude variations with a period of ~ 11 years in all these data (Belov 2000). Comparison with the sunspot numbers shows that the cosmic ray intensity is highest during sunspot minimum, then decreasing rapidly to a minimum value soon after “sunspot maximum”. As discussed in Sect. 5.7.1, this 11-year modulation is the consequence of the changing strength of the interplanetary magnetic field, and the degree of turbulence of the solar wind, throughout the solar cycle (McDonald 1998).

A prominent feature of both figures is the systematic difference between the recovery phases of the cosmic ray curves, from one sunspot cycle to the next (see also Fig. 5.7.5-1). Thus the recovery curves for 1959–1965 and 1981–1986 rise steadily in a monotonic manner to a relatively sharp peak coincident with sunspot minimum. By way of contrast, the recovery curves for 1951–1956, 1969–1976, and 1990–1997 rise fairly rapidly to a relatively “flat top” that continues for the remaining ~ 3 –4 years until the next sunspot minimum. This systematic difference between successive sunspot cycles is a consequence of the 22-year heliomagnetic cycle, as discussed in Sect. 5.7.5 (Jokipii 1991). The polarity of the Sun’s magnetic field reverses near the maximum of each sunspot cycle; by convention it is said to be of positive polarity when the solar magnetic dipole is pointed into the northern sky. The direction of the cosmic ray drift motions reverses when the polarity of the solar dipole changes, and the sign of the drift velocity, v_d , in the third term of the cosmic ray transport equation (Eq. 5.7.1-1) also changes. Using mathematical models to solve the cosmic ray transport equation, Jokipii and others

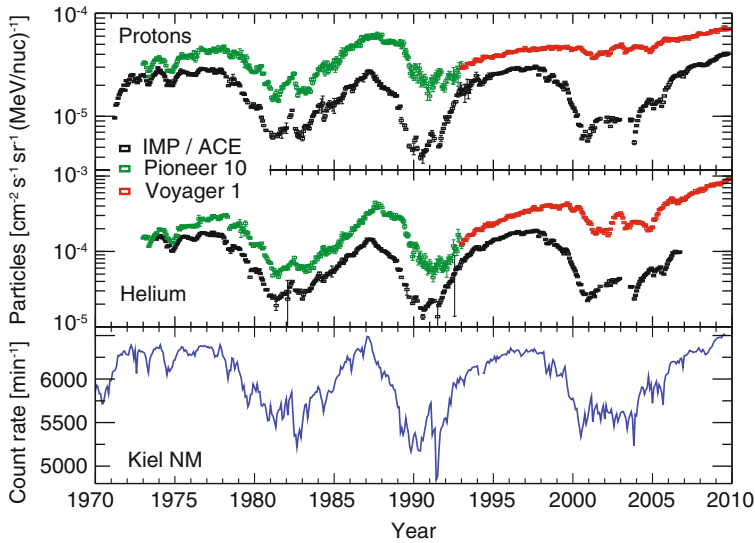


Fig. 7.2-2 The 11- and 22-year variations observed by satellite instruments, and the Kiel neutron monitor. The *top panel* shows the variation in the proton component of the galactic cosmic radiation while the *second panel* shows the helium component. Each component contributes about 50% of the production of the cosmogenic radionuclides observed at Earth, and their modulation effects are similar, but not identical. The IMP spacecrafts were orbiting Earth, while Pioneer 10 and Voyager 1 were steadily moving away from the Sun. Pioneer 10 had reached a distance of 40 AU from the Sun in 1991. Between 1991 and 2010 Voyager 1 moved from 50 to >100 AU. The Pioneer and Voyager data illustrate that the 11-year modulation effects are observed throughout the heliosphere. Further details regarding these data are given by McDonald (2000)

have shown that the alternate flat top and pointed top nature of the cosmic ray curve is a consequence of the reversal of the drift effects from one solar cycle to the next. (Jokipii 1991; Jokipii et al. 1977). A physical representation of the drift effects is given in Fig. 5.7.5-1. The cosmic ray transport equation indicates that the drift effect depends upon the product, qA , where q is the (signed) charge of the cosmic ray, and A is the (signed) magnitude of the Sun's polar dipole. As a consequence, the successive sunspot cycles are sometimes referred to as qA -positive (flat topped) and qA -negative (pointed top) cycles. For electron cosmic rays the sign of qA is reversed, and the flat topped cycles observed for the positive cosmic rays (H, He) become sharp topped, and vice versa.

Another prominent feature of Figs. 7.2-1 and 7.2-2 is the manner in which the cosmic ray intensity always returned to approximately the same value for each of the five sunspot minima – 1954 to 1996. This constancy in the instrumental record was originally used to question the validity of the cosmogenic data as a measurement of the intensity of the galactic cosmic rays. As discussed in the next section, it now appears likely that the similarity of the values at the sunspot minima between 1954 and 1996 was the result of five successive solar cycles of high solar activity.

The discussion of the modulation process in Sect. 5.7.3 has shown that the 11-year modulation is greater at low than at high energies. The >106 MeV IMP8 data shown in Fig. 7.2-1 are sensitive to lower energies than balloon and neutron monitor data (e.g. Fig. 6.5-1). That is, the measurements shown in Fig. 7.2-1 move to higher energies in the order IMP, balloon measurements, and the neutron monitors. The consequences of these two factors are clearly seen in Fig. 7.2-1. Note that for the solar cycle 1986–1997, the observed intensities decreased by $\sim 20\%$ (neutron), 30% stratospheric measurements, and 70% IMP. That is, the amplitude is greatest (70%) for the satellite detector that detects the lowest energy cosmic rays (>106 MeV), whereas it is smallest for the neutron monitors that only respond to >1 GeV cosmic rays. Note also that the amplitude in the Climax neutron monitor data is a factor of ~ 1.1 greater than for Moscow, the consequence of Climax being at a high altitude as discussed in Sect. 6.5.

So the question arises – will the amplitude of the 11-year (and other) variations in the cosmogenic data, for example, be similar to that of the neutron monitor (20%), the IMP detectors (70%), or some other value? The response functions in Fig. 6.5.2 provide a qualitative answer; the cosmogenic ^{10}Be curves lie between the IMP and the neutron monitor curves and therefore the 11-year amplitude will lie roughly midway between 20 and 70% . The inter-calibration curve in Fig. 6.6.1 gives a quantitative answer. It shows that variations in the ^{10}Be data will be ~ 2.3 – 2.5 times greater than those in the Climax neutron monitor data, in good agreement with the observations discussed in a later chapter.

The lowest curves shown in the top two panels of Fig. 7.2-2 are from the IMP spacecraft that remained in orbit around the Earth. Note that those data, and the Climax neutron data, returned to roughly the same intensities at each sunspot minimum. The other data shown in the top two panels are from the Pioneer 10 and Voyager 1 spacecraft that were moving steadily away from the Sun throughout this period. By the year 2000 the Voyager 1 spacecraft was >100 AU away from the Sun, and yet the 11-year modulation was still very pronounced at this great distance. The analysis of these and other data from distant parts of the heliosphere have provided strong validation of the cosmic ray propagation equation that is central to our understanding of the modulation process. It is these and similar spacecraft data that have established the validity of the “Rosetta Stone” approach to our interpretation of the cosmogenic data from the past.

7.3 The Long-term Variations

Anticipating the discussion of the cosmogenic data in Sect. 10.3, Fig. 7.3-1 shows a graph of two independent sets of ^{10}Be data from Greenland and the South Pole (McCracken et al. 2004). Note that these data are 22-year averages so that the 11- and 22-year cycles have been averaged out, leaving any longer-term variations in the data. It is clear that there are large ($\pm 25\%$) variations in both data sets, and that there is a high degree of correlation between the long-term (>50

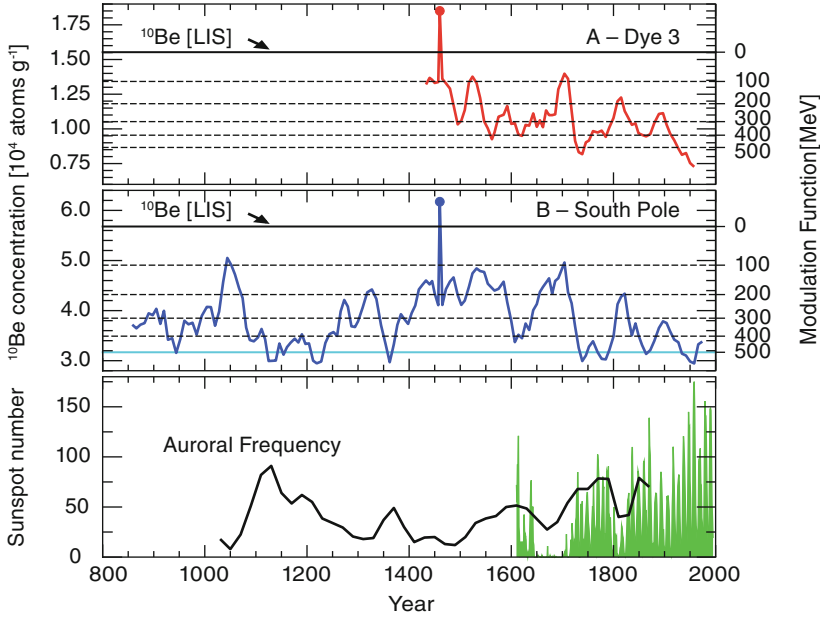


Fig. 7.3-1 The temporal variation of the cosmogenic ^{10}Be for the interval 850–1958 AD (McCracken et al. 2004), after compensation for the secular changes in the geomagnetic field. The data are the running 22-year averages. The ^{10}Be concentrations are given in atoms per gram. The *right-hand scale* and the *horizontal lines* show the modulation function (Sect. 5.7.3). The line labelled ^{10}Be (LIS) shows the ^{10}Be concentration corresponding to the cosmic ray local interstellar spectrum (LIS), as described in Sect. 5.7.2. The *bottom panel* presents the annual group sunspot number (1,600–2,000) and the normalized frequency of occurrence of the mid-latitude aurora between 1000 and 1900 AD (Krivsky and Pejml 1988). The two extreme points in the upper and the *middle panels* in the vicinity of 1460 AD are discussed in Sect. 18.3

year) variations. To further demonstrate this, the ~ 22 -year average data are plotted against each other (a “scatter plot”) in Fig. 7.3-2 (McCracken et al. 2004). While there are two noticeable outliers, it is clear that there is a close correlation between the ^{10}Be observed in the two polar caps (correlation coefficient = 0.87). The cosmogenic ^{14}C data prior to 1945 show variations in close agreement with these and, as we discuss in Chap. 17, this is a clear indication that the variations in both cosmogenic species are the result of long-term changes in the cosmic ray intensity at Earth.

These data are discussed in greater detail in Chap. 17; however, for the present we note

- The modulation function (right-hand scale) averaged over the solar cycle has been very low (~ 100 MeV) on three occasions that correspond to prolonged periods of very low solar activity as indicated by low sunspot numbers, and (further in the past) less frequent observation of aurora. That is, the cosmic

Fig. 7.3-2 Illustrating the similarity of the variations in the concentration of ^{10}Be in the northern and southern polar caps (McCracken et al. 2004) over a period of 525 years. The *scatter diagram* shows the plot of the 22-year average data from Dye 3, Greenland (*Y-axis*) versus South Pole (*X-axis*). The data are expressed as percentages relative to the averages for the period 1433–1958. The *solid red circle* represents the data for 1958, demonstrating that the production of ^{10}Be in the recent past has been one of the lowest since the Spoerer Minimum 1420–1530

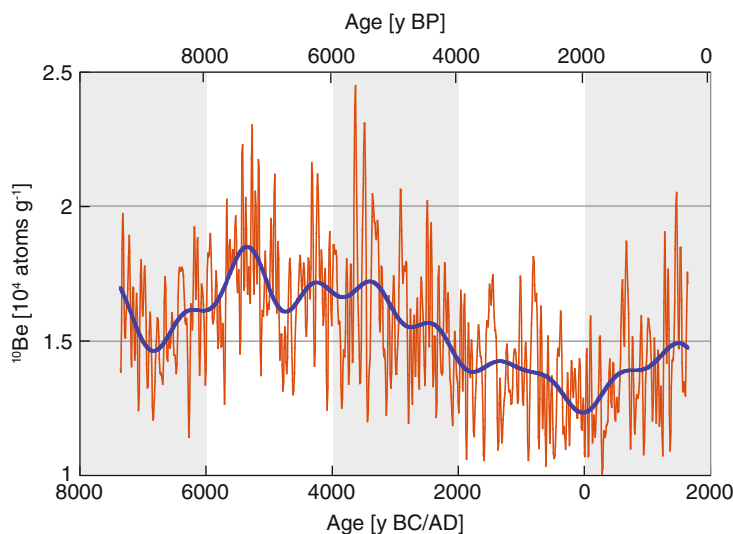
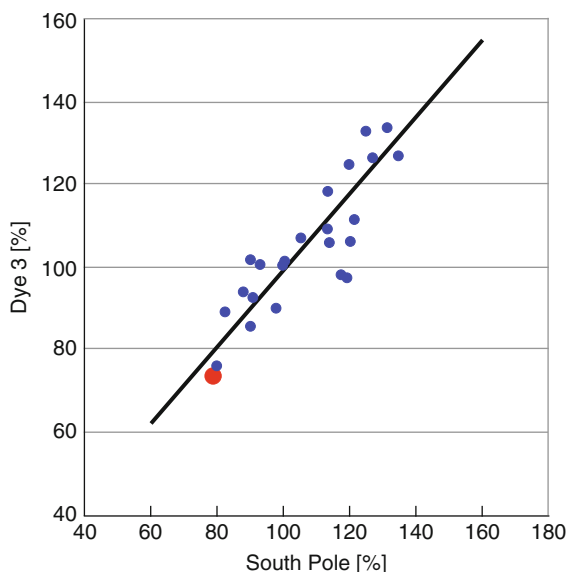


Fig. 7.3-3 The ^{10}Be paleo-cosmic ray record for the past ~10,000 years (the “Holocene”) low-pass filtered with 50- and 1,000-year filters, respectively. The long slow variation is largely due to changes in the Earth’s magnetic moment. The *sharp positive spikes* are the consequence of periods of low solar activity (Muscheler and Beer 2004)

ray spectrum near Earth has approached that prevailing outside the heliosphere (i.e. the local interstellar spectrum) during periods of very low solar activity.

- The cosmic ray intensity since 1950 has been one of the lowest in the last 1,150 years (middle panel).
- The long-term average intensity decreased rapidly between 1900 and 1950 AD, in agreement with the conclusions reached by Neher (Sect. 6.2). This decreasing phase between 1933 and 1954 can be seen in Fig. 6.2-3.

Figure 7.3-3 presents the ^{10}Be record from the GRIP ice core reflecting the cosmic ray intensity for the past almost 10,000 years. This is discussed in detail in Chap. 17; it is shown here to demonstrate the fact that the cosmic radiation at Earth has changed continually throughout this period. Without low-pass filtering and allowing for the 11-year variations, these data indicate that the production of ^{10}Be was at times (e.g. 5,000 years ago) ~ 3 times the production at other times (e.g. 2,300 years ago).

7.4 Forbush Decreases, Globally Merged Interaction Regions and Some Smaller Effects

The neutron monitor and IMP data contain some short duration variations that are unlikely to result in a recognizable change in the annual production rate of the cosmogenic radionuclides, which will however influence the ionization and radiation dose in the atmosphere. We briefly mention several of these for completeness.

The Forbush Decrease. The Forbush decrease appears as a large (up to 25%) decrease in the neutron monitor data (and also in other data records) that gradually decays over a period of a week or so (Fig. 7.4-1). A Forbush decrease is produced by the shock wave that is driven into the heliosphere ahead of a large CME (Cane 2000), and will typically extend over $>120^\circ$ of solar longitude (Sect. 5.7.6). The increased magnetic field strength in the shock provides an additional barrier to the cosmic rays reaching the region behind the shock – the shock literally sweeps cosmic rays outwards from the Sun. Although very noticeable in hourly or daily cosmic ray data, a very large Forbush decrease will only reduce the annual production rate of the cosmogenic radionuclides by $\leq 0.2\%$, which is undetectable. Depending on solar activity, there may be two or more Forbush decreases each month, with three or four large ones ($>15\%$ in neutron data) per year.

The cosmic radiation is usually quite anisotropic during the onset and first day or so – that is, the intensity depends quite strongly on direction. The wide scatter between the three curves during the first 2 days shown in Fig. 7.4-1 illustrates this, as the three detectors view the time-dependent cosmic ray fluxes from three directions spaced (roughly) 120° apart in (terrestrial) longitude as the earth rotates. These anisotropies provide a considerable amount of information regarding the approach, and extent of the shock wave that is responsible for the Forbush decrease. One of the primary goals of the “Spaceship Earth” network of neutron monitors (Sect. 6.3) is to be able to utilize these anisotropies as a “real time” monitor of the

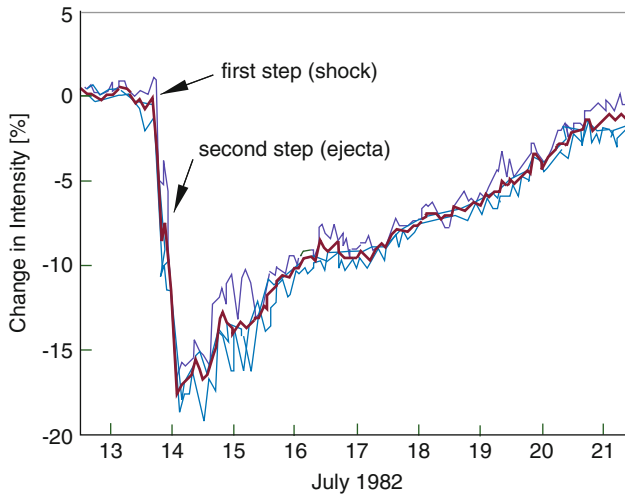


Fig. 7.4-1 The large Forbush decrease on 13 July, 1982 [after (Cane 2000)]. The figure shows the percentage decreases at three neutron monitors spaced about equally in longitude [Deep River (76°W), Mt. Wellington (147°E), Kerguelen Islands (69°E)], and the average shown by the *heavy line*. The first and second steps refer to the passage past the Earth of first the shock and then the ejecta shown in Fig. 5.7.6-1. This Forbush decrease occurred near the point of lowest intensity in the 11-year cycle and represents one of the lowest intensities that have occurred in the last 100 years. Note that the intensity recovered to the pre-event level after about 7 days, and compare this with the GMIR in Fig. 7.4-2

conditions in space within ~ 0.5 AU of Earth, something that we have no other way of investigating.

Global Merged Interaction Region (GMIR). This type of cosmic ray variation illustrates the complexity of the modulation process. From the 1950s, it was known that the neutron monitor counting rate would sometimes decrease in a step-like fashion as solar activity was rising to maximum. Unlike the Forbush decrease, the intensity then remained depressed for many months. Figure 7.2-1 illustrates this – note the 4% downward step in the Climax neutron monitor in 1956, and the two $>20\%$ downward steps in the IMP data in 1978 and 1979.

Figure 7.4-2 shows a GMIR as observed by IMP near Earth, and by the Voyager 2 spacecraft ~ 90 AU from the Sun, and almost all the way to the Termination Shock. Note that the cosmic ray intensity drops much more rapidly at Voyager 2 when the shock wave reaches it 8 months after it passed IMP. Examination of the data (and the interplanetary magnetic field data) shows that a series of separate, relatively small Forbush decreases was seen near Earth, leading to the steadily declining intensity stretching over a number of months. Then as the individual shock waves preceded outwards, the faster ones caught up with the slower ones, and between 30 and 60 AU they had merged into a single shock that extended around the 360° of solar longitude, hence the name – Globally Merged Interaction Region. Together, these merged shocks travelled outwards, presenting a major barrier that greatly reduced the cosmic radiation throughout the solar system.

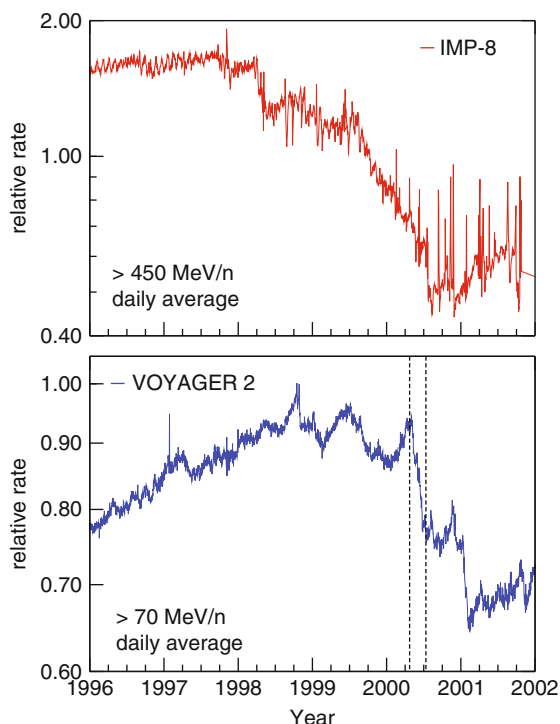


Fig. 7.4-2 A Global Merged Interaction Region (GMIR) observed throughout the solar system in 1999/2000 AD. *Top panel:* the “penetrating” counting rate (protons >450 MeV) at IMP 8 orbiting Earth. *Bottom panel:* >70 MeV/nucleon counting rate at Voyager 2, ~ 90 AU from the Sun. Strong solar activity had launched a series of shock waves over the full 360° of solar longitude that resulted in the intensity decline starting at IMP in September 1999. The first of the individual shock waves merged as they proceeded outwards, and a single globally merged shock wave went past Voyager 2 after ~ 8 months, resulting in a single, very large and fast (2 month) reduction indicated by the dotted lines. A second GMIR reached Voyager 2 about 6 months later

They do make a substantial change to the cosmogenic production rate (up to $\sim 10\%$ of the annual production) but this will be indistinguishable from the 11-year modulation.

Twenty-Seven Day Variations. In earlier years, before the advent of satellite instruments and the high counting rate IQSY neutron monitors, various statistical methods were used to show that the cosmic radiation tended to decrease by 1–3% for several days in a 27-day recurrent pattern that coincided with the well-known 27-day recurrence in geomagnetic disturbance. The 27-day periodicity made it clear that these variations were associated with the passage of a particular region of the Sun across the solar disc. The periodicity is now understood to be due to small Forbush decreases and the effects due to interplanetary shock waves produced when a somewhat faster solar wind ran into the back of a region of slower solar wind – the so-called co-rotating interaction regions.

The Diurnal Variation. As discussed in the introduction, this was the first type of time variation to be studied and established as a real effect. It is now known to be a small (0.5–2%) 24-h periodicity in the neutron monitor data. It is the direct consequence of the anisotropic flow of the cosmic rays inside the heliosphere. It is essentially a mean zero effect, and it definitely is too small to have any observable effect upon the cosmogenic data. However, its presence in the neutron monitor data provides day-to-day information on the spatial gradient of the cosmic radiation in the vicinity of Earth which is not provided in any other way. In particular, its direction in space exhibits a substantial 22-year (but not a 11-year) cycle, in which the cosmic rays reaching the Earth during a $qA > 0$ solar minimum (e.g. 1954) are stronger from a direction in the general direction of the Sun. During a $qA < 0$ solar minimum (e.g. 1986) the stronger flux is from a direction which is in the plane of the ecliptic, and 90° from the Sun (from the evening side of the Earth).

References

- Belov A (2000) Large scale modulation: view from the Earth. *Space Sci Rev* 93(1–2):79–105
- Cane HV (2000) Coronal mass ejections and forrush decreases. *Space Sci Rev* 93(1–2):55–77
- Forbush SE (1938) On world-wide changes in cosmic-ray intensity. *Phys Rev* 54(12):975–988
- Forbush SE (1946) Three unusual cosmic-ray increases possibly due to charged particles from the sun. *Phys Rev* 70(9–10):771–772
- Forbush SE (1954) World-wide cosmic-ray variations, 1937–1952. *J Geophys Res* 59(4):525–542
- Jokipii JR (1991) Variations of the cosmic-ray flux with time. In: Sonett CP, Giampapa HS, Mathews MS (eds) *The Sun in time*. University of Arizona Press, Tucson, AZ, pp 205–220
- Jokipii JR, Levy EH, Hubbard WB (1977) Effects of particle drift on cosmic-ray transport. 1. General properties, application to solar modulation. *Astrophys J* 213(3):861–868
- Krivsky L, Pejml K (1988) Solar activity, aurorae and climate in central Europe in the last 1000 years. *Publications of the Astronomical Institute of the Czechoslovak Academy of Sciences* (75):77–151
- McCracken KG, McDonald FB, Beer J, Raisbeck G, Yiou F (2004) A phenomenological study of the long-term cosmic ray modulation 850–1958 AD. *J Geophys Res-Space Phys* 109:A12103
- McDonald FB (1998) Cosmic-ray modulation in the heliosphere – a phenomenological study. *Space Sci Rev* 83(1–2):33–50
- McDonald FB (2000) Integration of neutron monitor data with spacecraft observations: a historical perspective. *Space Sci Rev* 93(1–2):263–284
- Muscheler R, Beer J (2004) Solar activity changes inferred from radionuclide records. *Geochim Cosmochim Acta* 68(11):A477

Chapter 8

The Solar Cosmic Radiation

8.1 Historical Overview

The first observations of the production of cosmic rays by the Sun were made in 1942 when ionization chambers in Europe, Greenland, the USA and New Zealand saw two short lived (~ 1 h), almost simultaneous, bursts of cosmic radiation (Forbush 1946). Three more bursts were seen by ionization chambers, Geiger counter telescopes, and neutron monitors up until 1956, and concurrent solar measurements indicated that they were usually closely correlated with the occurrence of a very large solar flare (Sect. 5.5) in the vicinity of a very large, complex sunspot group.

During the 1950s, it became clear that the solar cosmic radiation produced observable changes in the polar ionosphere as well. Thus long distance “forward scatter” communication circuits in the Arctic would suffer “polar cap absorption” (PCA) events following a major solar flare. Ionospheric opacity measurements (using an instrument called a “riometer”) and routine ionosondes (used since the 1930s in the study of long distance radio communications) were also found to be affected by solar cosmic rays. The latter were important – the archival records provided information back to the 1930s. Taken together, the ionization chamber and ionospheric data indicated that the production of cosmic rays by the Sun was a relatively rare phenomenon, and that the cosmic rays were accelerated in, or very close to, a solar flare itself.

The advent of the space age and the frequent observation of radio bursts from the sun changed this picture completely. By 1970, it was clear that the acceleration of ions and electrons was commonly associated with coronal mass ejections (CME) and solar flares. The electrons give rise to short-lived radio bursts; and 30–100 MeV ions are frequently observed by satellite borne detectors (Reames 1999). Much less frequently, >500 MeV cosmic rays are observed by neutron monitors on the surface of the Earth. The generation of >500 MeV solar cosmic rays is usually accompanied by the emission of X- and gamma rays, and intense radio emissions that may extend from 10 MHz to 80 GHz (Ryan et al. 2000).

The satellite observations soon showed that acceleration of solar cosmic radiation was often not occurring in the immediate vicinity of the solar flare, and sometimes there was no flare at all. With time, it became clear that much of the low-energy solar cosmic radiation seen by satellite detectors was being accelerated in the shock waves generated by coronal mass ejections (CME). The time dependence of the cosmic ray flux, its spectrum, its isotopic composition, and other properties are strongly dependent on where the cosmic rays are generated on the solar disk and upon the depth in the corona where they originate (Reames 1999).

Recently, it was proposed that the largest solar cosmic ray events have left a discernable signal in the ice deposited in the Arctic and Antarctic (Dreschhoff and Zeller 1994). This has provided the means to study the manner in which the production of solar cosmic rays has varied over the past 400 years (See, however, footnote on page 119). This shows that the frequency of occurrence of solar cosmic ray bursts has varied quite strongly from decade to decade in the past. Counter-intuitively, the studies have shown that the production of cosmic rays near the Sun shows an inverse correlation to overall solar activity. In other words, the frequency of occurrence in the past (e.g. near 1,900) was up to six times greater than that during the space age.

A number of names have been given to the production of high-energy ions and electrons by the Sun. The term “solar cosmic radiation” (SCR) is sometimes used to distinguish them from the galactic cosmic radiation (GCR). The most general name is the “solar energetic particle” (SEP) event, and this covers production at all particle energies, all nuclides, and electrons. An older term, the “solar proton event” (SPE), gained currency before the nature of the event was fully recognized, and is still used today, usually to describe the production of all the radiations listed above. A small fraction of all SEP events is observed by ground level detectors (e.g. neutron monitors), and they are frequently called “ground level events” (GLE).

8.2 The Observed Production of Cosmic Rays by the Sun

8.2.1 *Ground Level Events*

Starting with the first observation in February 1942, 70 GLE were recorded up to the end of 2009. They were observed by ionization chambers, Geiger counter telescopes, and a few neutron monitors up until 1956; by IGY neutron monitors until about 1966, and subsequently by IQSY neutron monitors and some remaining IGY monitors (see Chap. 6). There was a progressive increase in sensitivity going from one class of instrument to the next, and as a consequence the rate of observance prior to 1966 was less than it is now. It is usually accepted that about 15 GLE are observed per solar cycle, based on the worldwide IQSY neutron monitor network since 1966. While some GLE may be large (up to 5,000% above the pre-

event counting rate of a neutron monitor with a cut-off rigidity of 1 GV), 70% of them result in increases in counting rate that are less than 20%. Calculations show that it is only the largest that will result in production of cosmogenic radionuclides that is $>5\%$ of the annual production rate. Nevertheless, there is a finite probability that there may have been more than one in the past 10,000 years that resulted in $>100\%$ of the annual galactic cosmic ray production rate (see Sect. 8.3)

Figure 8.2.1-1 displays three examples of large GLE. The top panel displays the event of 25 July 1946 as seen by an ionization chamber in the USA. The second panel displays the most commonly referenced GLE that occurred on 23 February 1956. The third panel displays the large event of 20 January 2005. The ionization chamber was much less sensitive to solar cosmic rays than are neutron monitors; allowing for this lack of sensitivity the 15% increase in the top panel would correspond to an approximately fivefold increase if it had been observed by a high-latitude neutron monitor. Calculations summarized in Fig. 8.2.3-5 indicate that all three resulted in the production of small, but significant quantities of the cosmogenic radionuclides (Webber et al. 2007).

The three events in Fig. 8.2.1-1 exhibit quite different time profiles, and illustrate the complex nature of these events. Thus the 25 July 1946 GLE (panel a) lasted about twice as long as the other two. The parent solar flare occurred near the centre of the solar disk and, as a consequence, the GLE was followed about a day later by a very large Forbush decrease (Sect. 7.4). The GLE of 23 February 1956 (panel b) and 20 January 2005 (panel c) occurred near the western limb of the Sun, and as a consequence neither was followed by a Forbush decrease. All three GLE looked quite different when seen elsewhere on the Earth. Figure 8.2.1-1c shows this very well (McCracken et al. 2008). All four datasets (other than Climax) are from sea level neutron monitors with geomagnetic cut-offs well below 1 GV. The amplitudes of the GLE at all four locations would be the same if the solar cosmic radiation was isotropic (the same intensity from all directions in space). Note, however, that their amplitudes range from 5,200% (Terre Adelie in the Antarctic) to 300% (Nain in Canada). The amplitude at McMurdo was half that at Terre Adelie, only $\sim 1,300$ km away. These great differences show that the greatest intensities were streaming from a limited cone of directions in space, the intensity from all other directions being much smaller. This is not always so – a GLE from a parent flare near the centre of the solar disk (such as that in panel a) is much more isotropic, and the amplitudes may be roughly similar worldwide. The many differences from event to event, and place to place, are a consequence of the nature of the cosmic ray propagation between the Sun and the Earth, and beyond, along the Parker spiral field (Fig. 5.5-4). In general, it is important to note that these differences exist when using GLE data for environmental applications. Estimates or conclusions based on a few datasets may result in serious misinterpretation.

As we will discuss in Sect. 8.3.1, the spectrum of the solar cosmic radiation is much steeper than that of the galactic cosmic radiation. In other words, the intensity falls off steeply towards higher energies, and only three GLE in the past 67 years resulted in measureable intensities near the equator ($E > 10\text{--}15$ GeV/nucleon).

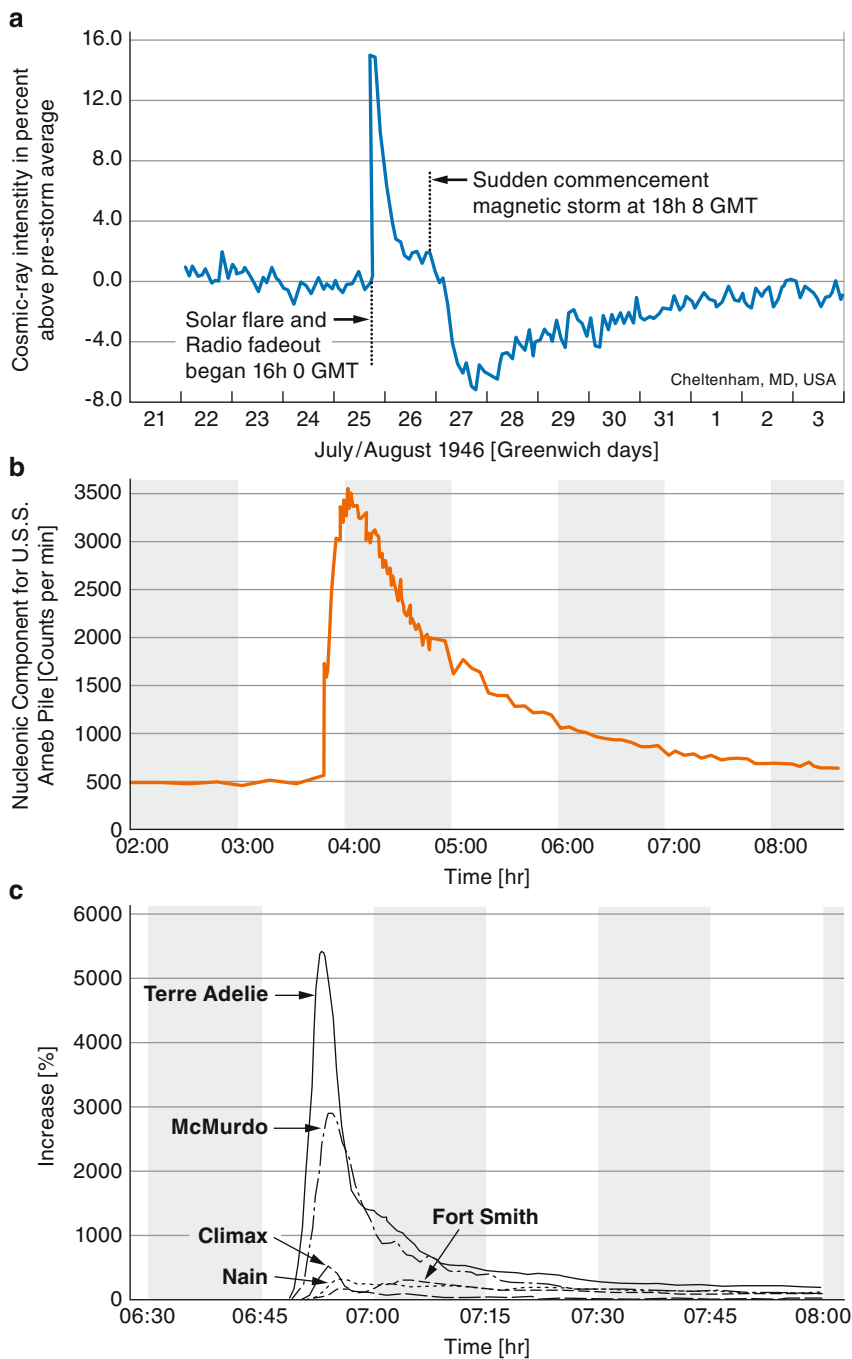


Fig. 8.2.1-1 Three large ground level events seen by ground level cosmic ray detectors. (a) 25 July 1946, as observed in Cheltenham, Maryland by an ionization chamber (Forbush 1946), (b) 23 February 1956, observed in Wellington, New Zealand, by a neutron monitor (Meyer et al. 1956), and (c) 20 January 2005, as observed by five neutron monitors (McCracken et al. 2008)

8.2.2 SEP Events Observed by Satellites

Above the Earth's atmosphere, and outside the geomagnetic field, a satellite borne detector sees an entirely different world compared to that seen by a ground level detector. SEP events are very common; and while the majority of GLE exhibit an intensity increase of less than a factor of three, four orders of magnitude increases are common in satellite instruments measuring ~ 10 MeV cosmic rays. This is a very complex subject and beyond the scope of this book, and only a brief outline is given here. For a detailed discussion, the reader is directed to the comprehensive review by Reames (Reames 1999).

Figure 8.2.2-1 illustrates the two main classes of SEP events discussed in the literature: the gradual event (panel a), and the impulsive “ ^3He -rich” event (panel b). Figure 8.2.2-2 provides an example of a “classical” gradual event. Note that in both the cases the event lasts for a number of days.

Figure 8.2.2-3 is a schematic diagram of the heliomagnetic fields at the time of a gradual event. A CME has been ejected from the Sun, and is driving a shock wave ahead of it. Ions and electrons are accelerated in the outward sweeping shock wave and some have then escaped from the shock and are propagating outwards to the vicinity of Earth. Others have escaped into the magnetic fields behind the shock. In front of the shock, the intensity increases steadily at low energies as more and more ions are accelerated as the shock moves outward.

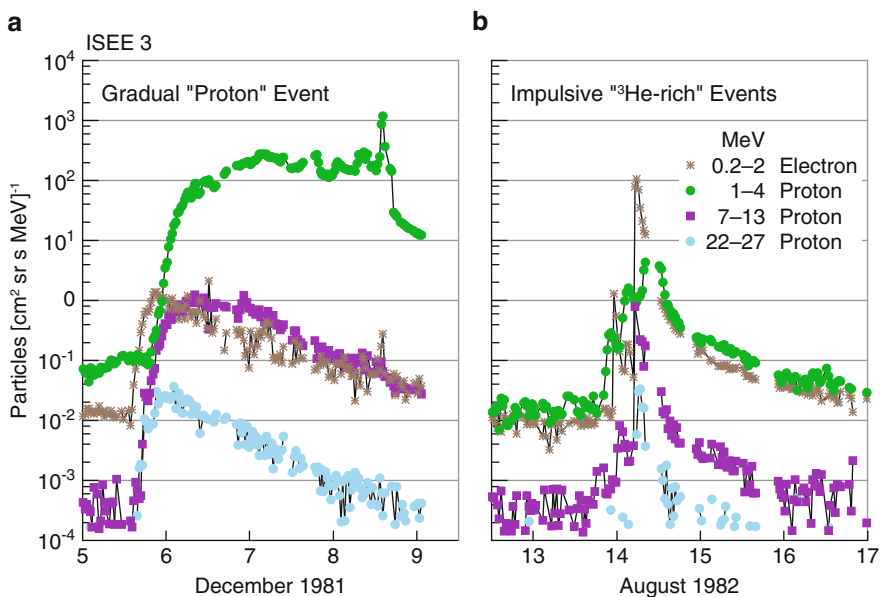


Fig. 8.2.2-1 Intensity versus time profiles of electrons and protons in (a) gradual and (b) impulsive SEP events (Reames 1999). The gradual event was due to a CME without an impulsive flare. The impulsive events came from a series of flares with no observed CMEs

The time profile of the gradual SEP depends strongly on where the CME is located on the solar disk relative to the observer, as illustrated by Fig. 8.2.2-3. For the example in the left hand panel, the CME erupted 53° west of the central solar meridian; the intensity peaked within 18 h and then fell relatively rapidly. The CME was 45° east of the central meridian in the example in the right hand panel, and the intensity did not reach its maximum for 3 days and then decreased slowly thereafter. We note that “gradual” refers to the overall duration of the SEP event, and does not necessarily refer to the rise time.

The particle intensity frequently increases by a factor of ~ 10 as the shock wave sweeps past the detector (Figs. 8.2.2-2 and 8.2.2-3). This is due to accelerated particles that are still trapped in the strong magnetic fields in the shock wave. After the shock wave has gone by, the detector samples the solar particles that have been trapped in the magnetic fields being swept outwards by the CME.

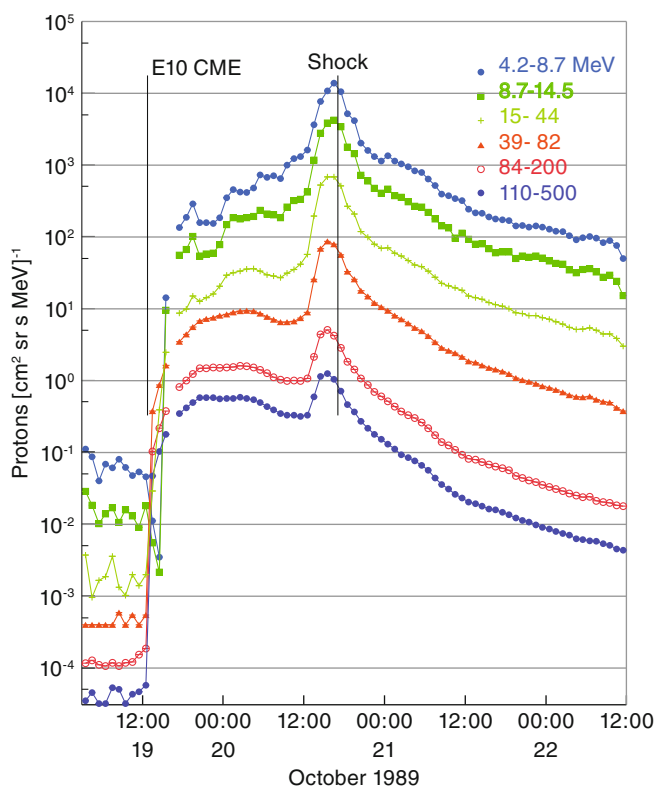


Fig. 8.2.2-2 The “classical” gradual SEP event of 19 October 1989 (Reames 1999). A CME occurred 10° to the East of the central solar meridian and launched an interplanetary shock wave that reached Earth a day later. The intensities at lower energies rose by up to an order of magnitude as the shock approached Earth. They increased by another order of magnitude at all energies as the shock passed Earth. A Forbush decrease started in the neutron monitor record as the shock passed by the Earth

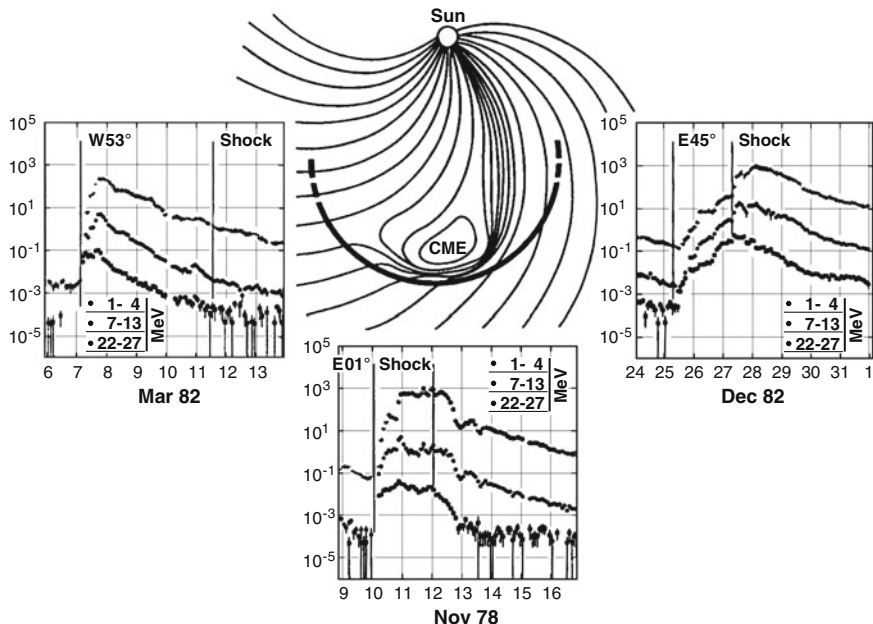


Fig. 8.2.2-3 Illustrating the “interplanetary geometry” of the gradual SEP event (Reames 1999). *Top* = 1–4 MeV; *Middle* = 7–13 MeV; *Bottom* = 22–27 MeV. The intensity–time profiles are shown for three events from different solar longitudes relative to the CME and the shock. When the CME was 53° west (*left hand profile*), the event rose and decayed relatively rapidly. Contrast this to the *right hand profile* (CME 45° East) where the event rises and falls slowly and the event persists over many days. The time of occurrence, the location of the CME on the solar disk, and the arrival of the shock near Earth are shown

The impulsive SEP event (Fig. 8.2.2-1b) has quite different characteristics from those of the gradual event. The intensity rises rapidly to a maximum, and then decays away more rapidly than in the case of the gradual event. Its most striking feature, however, is the high abundance of the relatively rare isotope, ^3He , in the SEP flux (Reames 1999). Thus the ratio of ^3He to ^4He is frequently 100 or more times greater than that in the solar corona, or in the gradual events. It is thought that the impulsive event represents acceleration relatively deep in the corona, in close proximity to a flare, and that the high abundance of ^3He is due to a resonant process wherein the ^3He is selectively accelerated compared to ^4He .

As in the case of the GLE, the interplanetary magnetic field has a strong influence on the SEP particles. This can be seen in Fig. 8.2.2-4 which shows that the gradual events (left panel) are seen after solar activity on nearly the whole visible disk of the sun. This is explained in terms of the acceleration taking place in a CME, which can spread out 60° on either side of the location of the solar activity (e.g. Fig. 8.2.2-3). The impulsive event, however, is only seen from a limited range of solar longitudes. This is explained in terms of these impulsive event cosmic rays

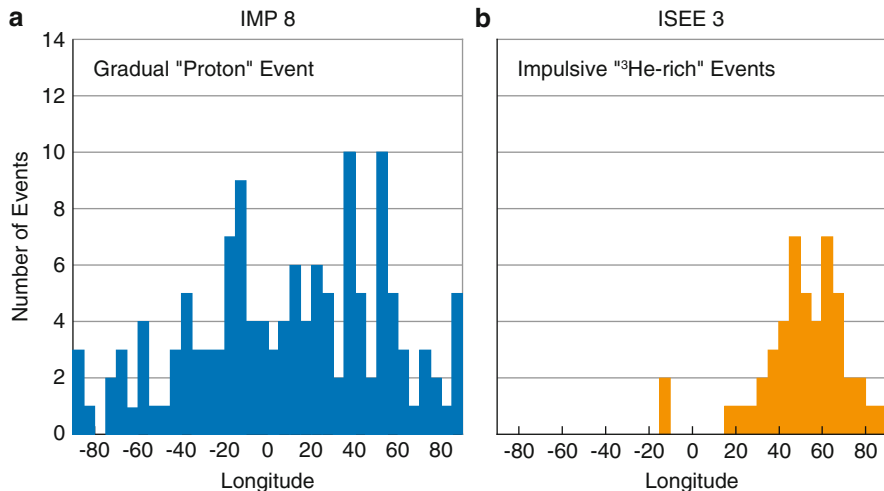


Fig. 8.2.2-4 The source regions on the Sun for the SEP events seen at Earth (Reames 1999). *Left panel:* The longitudes of the solar events corresponding to the large, gradual SEP events. *Right panel:* The source longitudes for the impulsive ^3He rich events. There is a similar effect for GLE; the fastest rising, shortest lived GLE being associated with a range of solar longitudes similar to that in *panel b*, while slower rising, longer lived GLE have been seen to the full extent of the range given in *panel a*. Solar longitude is quoted relative to the central meridian of the Sun as seen from Earth

being generated in a relatively small area of the Sun, the interplanetary magnetic field then inhibiting them from spreading out very far in solar longitude (see Sect. 8.3.2). In other words, from the western portion of the Sun as seen from Earth, the SEP particles are guided to the general vicinity of the Earth by the interplanetary magnetic field.

The cosmic ray “fluence” is commonly used in discussing the practical implications of SEP events. It is defined as the time integral of the integral intensity (above a specified threshold E_t) over the whole event

$$F(E_t) = \int J(E_t, t) dt$$

It is usually quoted as the integral over a solid angle of 4π , in units of particles cm^{-2} (referred to as the omnidirectional fluence). It is frequently quoted as the >30 MeV fluence (i.e. the integral of all particles above 30 MeV) or the >10 MeV fluence. Clearly, it is an appropriate quantity in estimating the radiation dose, the amount of ionization, or the cosmogenic nuclei production in the polar caps during a SEP event.

The proton flux unit (PFU) is frequently used in the discussion of space weather and the practical implications of SEP events. It is a directional flux, measured, in the units of particles $\text{s}^{-1} \text{cm}^{-2} \text{steradians}^{-1}$.

8.2.3 *Paleo-Cosmic Ray Measurements of SEP Events*

Until 2000 AD, there were no well-accepted physical records of the occurrence of SEP events prior to the commencement of instrumental measurements. In 1962, Lal and Peters predicted that large GLE such as that in 23 February 1956 would result in significant production of the cosmogenic radionuclides (Lal and Peters 1962). As with most of their pioneering work, they were right; however, it took 50 years to demonstrate it. There was a suggestion that solar events were recognizable in the ^{14}C record circa 1890 and an association of thin layers of nitrate in polar ice cores with SEP; however, neither association was widely accepted.

In the 1990s, the “Rosetta Stone” approach was applied to this problem (Vitt and Jackman 1996; McCracken et al. 2001). It involved the use of the SEP fluxes observed by satellites during a number of large SEP events over the previous 10 years, known atmospheric chemical processes, and a detailed mathematical model that simulated worldwide atmospheric transport processes. Together, they were used to calculate the origin of all the nitrate reaching the polar caps (Vitt and Jackman 1996). The calculated production was found to be consistent with the nitrate concentrations in the thin (~6–12 weeks duration of deposition) layers of nitrate observed in ice cores obtained in the polar caps. There also was a good correlation between the thin nitrate layers in the ice cores, and the largest GLE, and satellite observed SEP events between 1946 and 1972 (see below). Figure 8.2.3-1 presents the nitrate record obtained at “Summit”, a high-altitude drilling location in the centre of Greenland. The cores were cut into thin samples – about 15 samples per year – and dated using the known dates of volcanoes in Iceland (that produce very clear layers containing dust and sulphuric acid) in the ice core. The “discovery” event was the very large, short-lived increase in the ice layers from August to September 1859; that corresponded well with the Carrington–Hodgson white light flare of 1 September 1859 (McCracken et al. 2001) (see Sect. 5.5). The SEP event is superimposed on an annual variation in nitrate due to the seasonal transport of nitrate generated by photochemical processes initiated by solar ultra-violet radiation in the equatorial regions (this solar UV is continuously produced by the Sun). The electrical conductivity responds to both the sulphuric acid from volcanoes and the acidic NO_3 ions produced by the solar cosmic rays. Thus the SEP event in Fig. 8.2.3-1 is seen in both the NO_3 and the conductivity graphs, while the Mt. Hekla volcano (Fig. 8.2.3-2, panel a) was only seen in the conductivity trace.¹

¹ As this book was in preparation for printing, results from a number of new ice cores failed to provide independent verification of the “Carrington event”. Support for the use of nitrate events as proxies for SEP has been provided, however, by new evidence that two large nitrate events were observed in both the Arctic and Antarctic at the time of large GLE (e.g., July 1946, Fig. 8.2.1-1 and 8.2.3-3) and detailed investigations of these are in progress. We still provide the nitrate results in Section 8.2.3, while warning the reader that some or all of them may require revision in the next year or two.

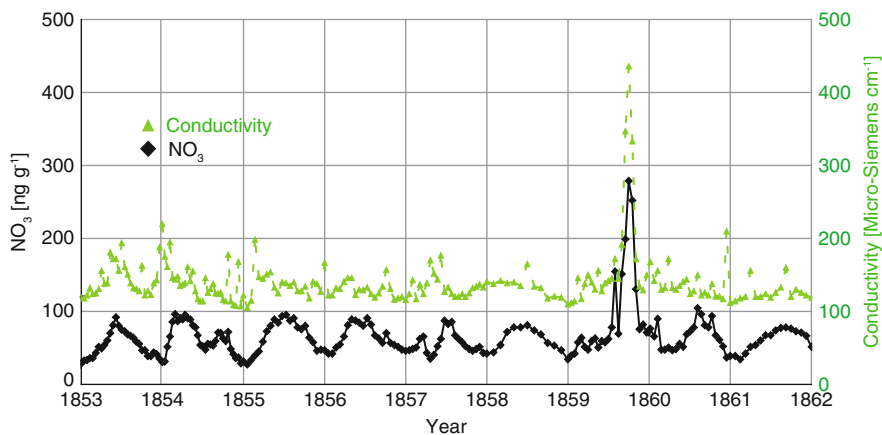


Fig. 8.2.3-1 The impulsive nitrate event of September 1859 is the largest in the record for the interval 1562–1950 and occurred in close correlation with the “white light flare” observed by Carrington and Hodgson on 1 September 1859 (McCracken et al. 2001). The nitrate data are in units of nanograms per gram of ice. This event was computed to have a >30 MeV fluence of $18.8 \times 10^9 \text{ cm}^{-2}$. The *green curve* gives the conductivity of the melted ice in $\mu\text{S cm}^{-1}$; this provides “time markers” from the occurrence of volcanic activity in Iceland as well as responding to the acidic NO_3 ions produced by the solar cosmic rays (See footnote, page 119)

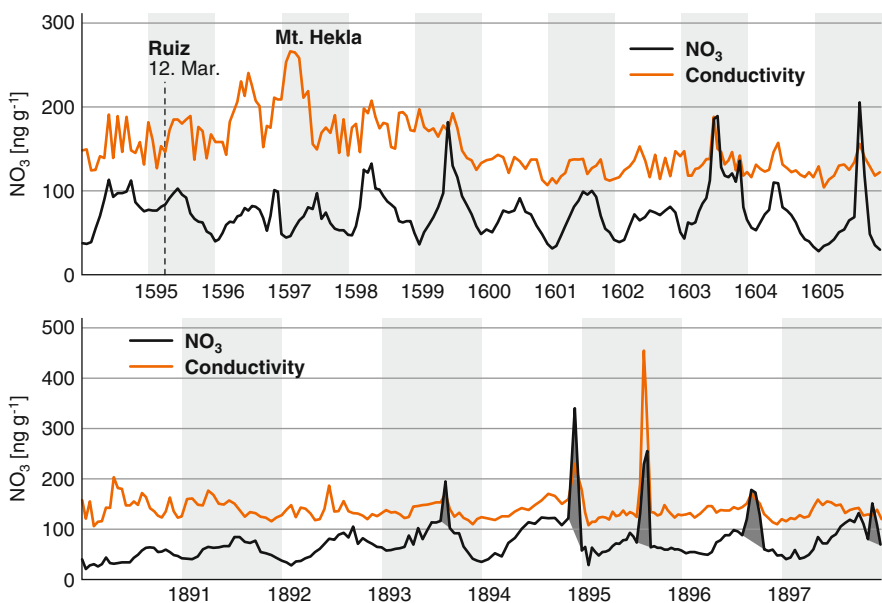


Fig. 8.2.3-2 Two intervals showing frequent occurrence of impulsive nitrate events (McCracken et al. 2001). (a) The interval 1594–1606 with frequent events during a period of high solar activity as evidenced by the frequent observation of mid-latitude aurora, and (b) the interval 1890–1898 with five very large events

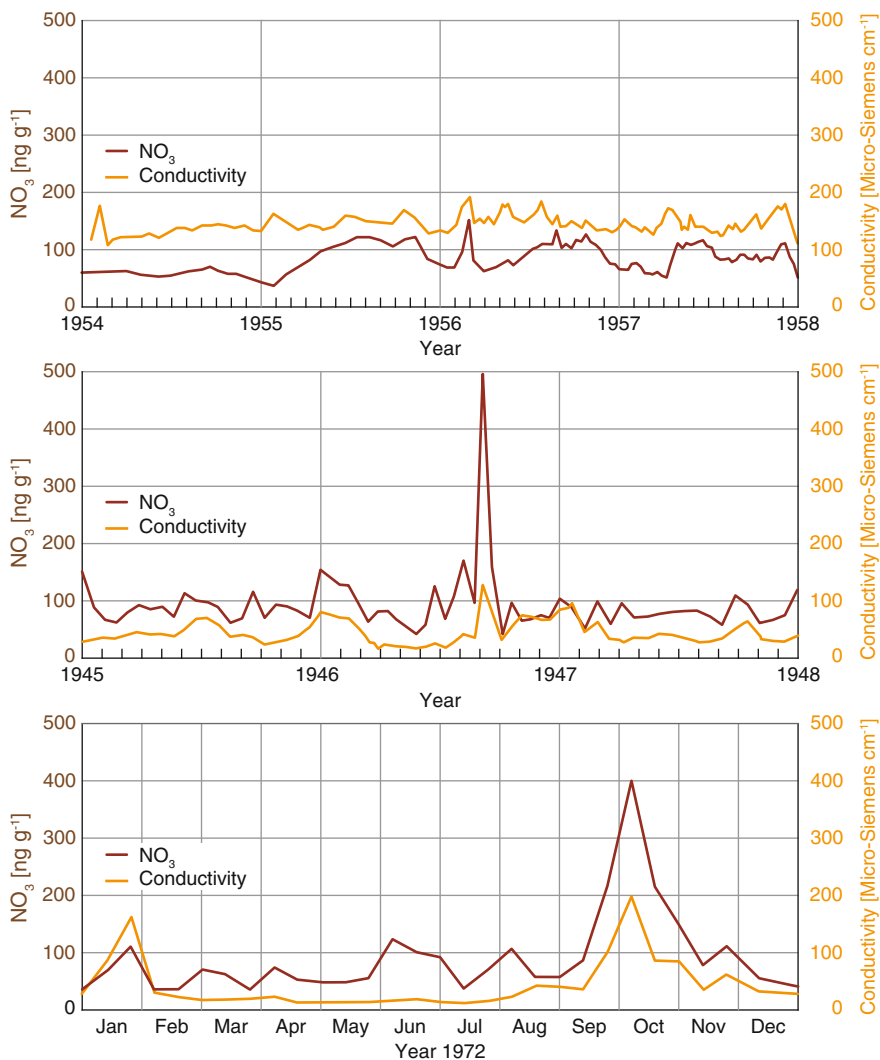


Fig. 8.2.3-3 Three impulsive nitrate events associated with the observation of solar cosmic rays with ground level and satellite borne detectors. (Panels *a* and *b*) The nitrate data coincident with the large GLE of 23 February 1956, and 25 July 1946 (Fig. 8.2.1-1), and (c) coincident with the large gradual SEP events of August 1972, observed with satellite borne instruments (McCracken et al. 2001). (See footnote, page 119)

Figure 8.2.3-2 presents the nitrate results for two interesting periods that illustrate very frequent production of solar cosmic rays. The bottom panel shows that five of the largest nitrate events observed since 1572 occurred in the sunspot cycle between 1890 and 1901. While this was a relatively weak sunspot cycle, the excellent geomagnetic records for that time show that very large magnetic storms

were very frequent as well. We will discuss this period again in Sect. 17. The top panel is from the period immediately prior to Galileo's observation of sunspots in 1609. The nitrate peaks show SEP events were fairly frequent and contemporary records show that mid-latitude aurora were seen frequently (McCracken et al. 2001). This was a period of intense solar activity.

By way of sharp contrast, only two nitrate events were observed for the 50-year interval 1650–1700, this being during the Maunder Minimum of solar activity.

Further examples of the nitrate enhancements are given in Fig. 8.2.3-3. This shows the nitrate enhancements at the time of three very large solar cosmic ray events observed by ground level and satellite borne detectors. The top two diagrams correspond to the very large GLE in February 1956 (Fig. 8.2.1-1b) and July 1946 (Fig. 8.2.1-1a). The third corresponds to a series of large gradual SEP events seen by satellite detectors in August 1972. These three events and the “Rosetta Stone” calculations have provided the crucial ability to convert the observed increases in the nitrate concentration from past centuries to the fluence measurements of our modern space era (McCracken et al. 2001).

In passing, we note that the amplitude of a SEP event in the nitrate record is strongly dependent on season as a consequence of atmospheric factors. Thus they are larger in winter than in summer, and hence may be large in the data from one polar cap, and negligible in that from the other. It has been estimated that up to 25% of large SEP may have missed detection for this reason (McCracken et al. 2001). Unfortunately, there is only one centennial scale record at the time of writing that extends from 1572 to 1950. Clearly more records are desirable.

As discussed above, the top panel of Fig. 8.2.3-2 shows that there were five very large SEP events in the interval 1892–1898. Examination of the cosmogenic ^{10}Be for that interval and neutron measurements made in the troposphere and stratosphere during the GLE of 23 February 1956 (Fig. 8.2.1-1b) led to the conclusion that about 5–10% of the annual (global average) ^{10}Be production rate for those years was due to the SEP events detected by the nitrate record, and that other large SEP would also produce increases of that magnitude in the cosmogenic record.

The “Rosetta Stone” approach has confirmed this experimental deduction. Using the observed or estimated cosmic ray spectra from 13 large SEP events (as given in Fig. 8.3.1-2), and the known production rate of ^{10}Be and ^{36}Cl as a function of energy (Fig. 8.2.3-4), the annual production of cosmogenic ^{10}Be and ^{36}Cl by both the galactic and solar cosmic radiation was estimated for the interval 1939–2005. Figure 8.2.3-5a, b displays the results. The steep nature of the SEP spectra (with little flux >1 GV) means that nearly all of the cosmogenic production occurs in the polar caps (i.e. low cut-off rigidities), with the majority being generated in the stratosphere. The estimates are plotted for the two extreme assumptions: (a) that all of the cosmogenic radionuclides are precipitated in the polar caps, (b) that there is complete atmospheric mixing prior to deposition in the polar caps thereby reducing the amount of the cosmogenic radionuclides reaching the polar caps (global average).

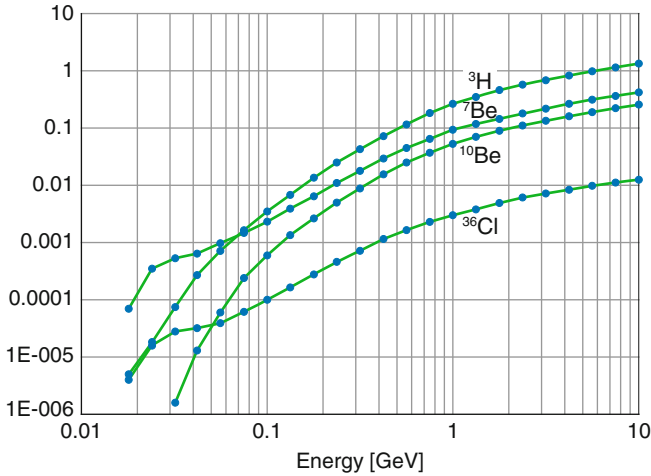


Fig. 8.2.3-4 The specific yield functions for the production of several cosmogenic radionuclides extended to the low energies where the solar cosmic ray intensities are the greatest (Webber et al. 2007). Note the high values of $S(E)$ for $30 < E < 70$ MeV for ${}^7\text{Be}$ and ${}^{36}\text{Cl}$ as a consequence of the high values of their excitation functions at low energies as displayed in Figs. 10.3.2-1 and 10.3.2-2

Consideration of the Fig. 8.2.3-5a, b and comparison with the spectra in Fig. 8.3.1-2 indicates the following:

- The SEP event with the “hardest” spectrum with high intensities in the vicinity of 1 GeV (spectrum 1–23 February, 1956) is the most prominent of all the events in the computed ${}^{10}\text{Be}$ record. In passing we note that this GLE resulted in a 50% increase in the counting rate of the Huancayo neutron monitor near the equator.
- The large SEP events with softer spectra (order of magnitude higher intensities at low energies but much less than (1) at high) are the most prominent in the ${}^{36}\text{Cl}$ plot (spectra 3 and 4). This is a consequence of the very high production cross-sections in the range 10–30 MeV (see Sect. 10.3 and Fig. 10.3.2-2) that results in the “bump” in the specific yield function in Fig. 8.2.3-4 in the range 30–70 MeV.
- The specific yield function for ${}^7\text{Be}$ has a low-energy enhancement similar to ${}^{36}\text{Cl}$, and, by inference, will be sensitive to the “soft” spectrum SEP similar to ${}^{36}\text{Cl}$ in Fig. 8.2.3-5b.
- The majority of the increases in the ${}^{10}\text{Be}$ due to SEP events will be “lost” in the 11-year variation, particularly in the case of global averaging
- The largest events occurring near the beginning or end of the 11 year cycle may not be confused with the 11 year variation, particularly in the case of ${}^{36}\text{Cl}$ and ${}^7\text{Be}$.

Clearly, the detection of SEP events in the cosmogenic record is marginal at best. However, it is likely that more accurate measurements, the summation of several independent records, or larger or multiple events in the past, would result in statistically significant cosmogenic records of SEP events.

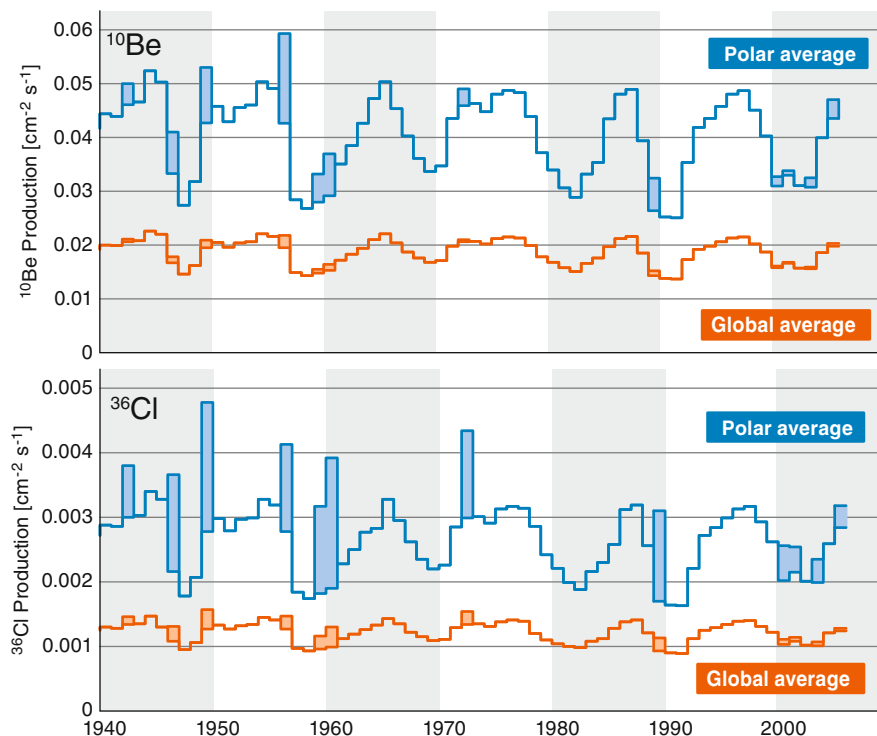


Fig. 8.2.3-5 The computed production of ^{10}Be and ^{36}Cl by both the galactic and solar cosmic radiation since 1939 (Webber et al. 2007). *Top panel:* ^{10}Be . *Bottom panel:* ^{36}Cl . The *top curve* in each panel assumes no latitudinal mixing (see Chap. 13) while the *bottom* assumes complete global mixing prior to precipitation. The reality is intermediate between these two limits, but close to complete mixing due to the majority of cosmogenic radionuclides being produced in the well-mixed stratosphere (Chap. 13)

The cosmogenic record is being used increasingly to estimate environmental and other changes that will occur in the future. For example, the ^{10}Be record has been used to estimate the total solar irradiance (TSI) in the past (Sect. 17.5) in the estimation of climate change in the future (IPCC 2007). The modulation function, Φ , will be increasingly used for archaeology, paleo-climate and other applications to quantify solar activity in the past. It is noted that the presence of undetected SEP events will introduce errors into any of these applications. Thus the modulation function Φ and the estimates of the interplanetary magnetic field assume that there is no SEP contribution. The nitrate record permits partial correction for this effect, however, as discussed above, it is estimated that only 75% of the large SEP events have been detected to date. Further, the evidence suggests that SEP events are more frequent at times of low solar activity, such as in the vicinity of the Grand Minima (McCracken et al. 2004). The GCR intensity is already high at such times, and undetected SEP events will further increase cosmogenic production, leading to an underestimate of Φ .

The enhanced sensitivity of ^{36}Cl to solar cosmic rays could provide a means to detect and allow for such events.

8.3 Overall Characteristics of the Solar Cosmic Radiation

Based on ground level, satellite, and paleo-cosmic ray observations, the characteristics of SEP events can be summarized as follows. As indicated in the discussion of both the GLE, and satellite data, there is a wide spread in the characteristics from one event to another (Reames 1999). This spread is understood to be due to event to event differences in: (1) the acceleration processes; (2) the location of the point of acceleration on the solar disk (see below); (3) the scattering characteristics of the heliospheric magnetic field between the Sun and the Earth; (4) whether there has been an earlier shock wave from a similar location on the solar disk; and other factors.

8.3.1 The Energy Spectra

Figure 8.3.1-1 displays the data from a relatively large SEP event that was observed both by neutron monitors and by satellites in 1989. It was about one-tenth as intense as the largest GLE observed in the past 70 years. The figure shows the extreme differences between the spectra of the galactic and solar cosmic rays, e.g. at 100 MeV/nucleon the solar proton intensity was five orders of magnitude greater than the galactic. The rapid increase in intensity towards low energies compensates for the decreasing yield functions in Fig. 8.2.3-4; so significant cosmogenic production occurs at energies as low as 20 MeV/nucleon very high in the stratosphere. For energies ≥ 1 GeV (i.e. that yield significant radiation doses at aircraft altitudes, or that are efficient generators of cosmogenic radionuclides), the energy spectrum was much steeper than the GCR (E^{-6} , compared to $E^{-2.5}$).

We note in passing that it is common practice to quote solar cosmic ray intensities in terms of particles per square centimetre, while galactic cosmic ray intensities are usually given per square metre. Thus the spectra in this chapter are quoted per square centimetre, and the GCR spectrum in Fig. 8.3.1-1 has been converted to those units (it is quoted in Fig. 5.7.3-1 per square metre).

The spectra vary considerably from one event to another, as summarized by the integral spectra in Fig. 8.3.1-2. To some extent, this is due to where the solar cosmic radiation originated on the solar disk (see next section). The different energy sensitivities of the various detection methods, and these differences in spectral shape suggest that the various methods will not respond to all SEP events in a similar manner. Thus the SEP events observed by one detection method may not be seen by another, and vice versa.

Fig. 8.3.1-1 The differential energy spectra of H and He in the 29 September 1989 event (Lovell et al. 1998). The data points in the *upper left hand corner* are from three satellites, GOES 6 and 7 and IMP8, while the *shaded region* is based on neutron monitor data. This was one of the largest GLE in the past 75 years. The *fitted curves* are those predicted by the shock theory of Ellison and Ramaty (Ellison and Ramaty 1985). The GCR proton spectrum is for a modulation function of 600 MeV, as given in Fig. 5.7.3-1

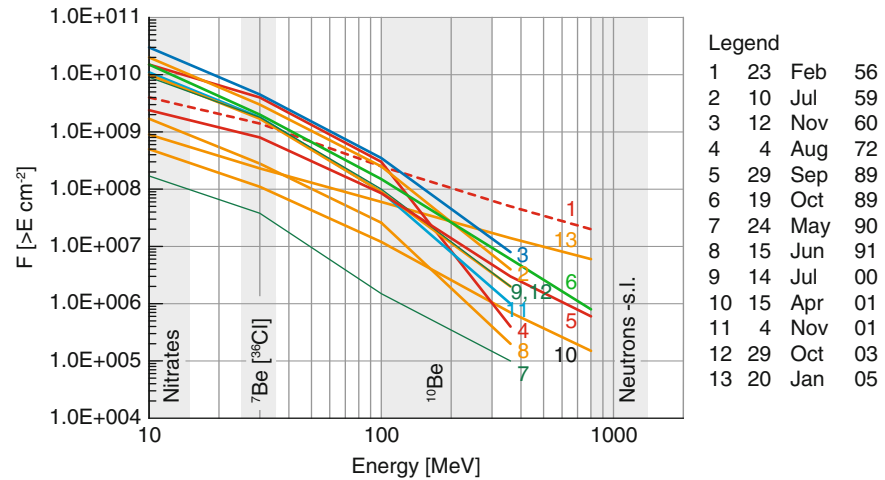
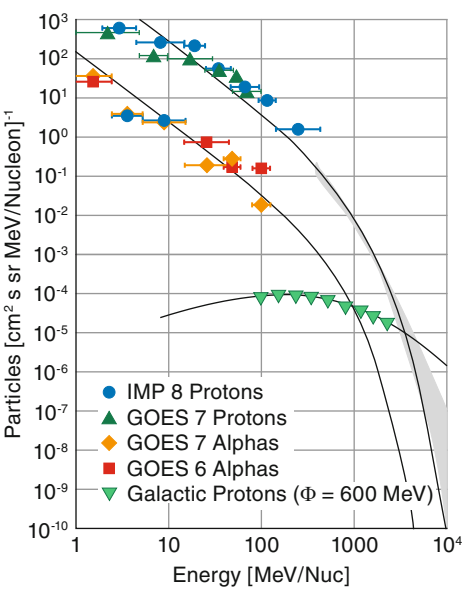


Fig. 8.3.1-2 Illustrating the variability of the integral energy spectra of thirteen SEP events (Webber et al. 2007). Some [curve 1 (23 Feb 1956) and curve 13 (20 Jan 2005)] extend to high energies. There is a larger class [e.g. events 3 (12 Nov 1960) and 6 (19 Oct 1989)] which had high intensities at low energies, with the spectrum dropping rapidly above 100 MeV/nucleon. The *shaded bands* indicate the energies of greatest sensitivity of the various detection methods

8.3.2 *The Effect of Longitude Relative to the Central Solar Meridian*

As discussed in Sect. 5.5, the outward flowing solar wind results in the lines of force of the heliomagnetic field being in the shape of Archimedes spirals (Fig. 5.5-4) (Parker 1958, 1963). When injected into that field near the Sun, the solar cosmic rays spiral outwards along the lines of force. As a consequence, solar cosmic rays are preferentially delivered to the Earth if they are generated on the western portion of the solar disk. (West, here, refers to the western portion of the Sun as seen from the Earth. The rotation of the Sun, in this convention, is from East to West, so the sunspots are moving to the west from day to day, and ultimately disappear over the western limb of the Sun.) The strength of the heliospheric magnetic field decreases rapidly between the Sun and the Earth, resulting in “pitch angle focusing” wherein the pitch angle, θ (the angle between the cosmic ray velocity vector, and the magnetic field vector) obeys the relationship $(\sin^2\theta)/B = \text{constant}$). Since B decreases by about four orders of magnitude between the Sun and Earth, θ can become quite small; however the pitch angle focusing is opposed by the random scattering effects of the irregularities in the heliospheric field (Sect. 5.5). As a result of these two opposing effects, the solar cosmic rays will soon be travelling approximately parallel to the interplanetary magnetic field with a distribution of θ about that direction that will vary from event to event.

As a result of the above, the cosmic rays from the western disk arrive at Earth with little delay since they are travelling (roughly) parallel to the Archimedes spiral. Thus the first 1 GeV particles reach the Earth about 11 min after they leave the vicinity of the Sun, and may be all moving approximately parallel to the lines of force of the heliospheric magnetic field. For 100 MeV particles the time of flight is ≥ 30 min. At all energies, they may arrive at Earth as a highly anisotropic beam of radiation (as discussed in connection with the GLE of 20 January 2005, Sect. 8.2.1). In that case, the intensity of ~ 1 GeV particles went from zero to maximum in a period of ~ 6 min for neutron monitors so situated that they saw the highly anisotropic beam of particles when it first reached Earth. The GLEs of 23 February 1956 and 20 January 2005 in Fig. 8.2.1-1 are good examples of this class of event (McCracken et al. 2008), as are the prompt events at low energies (Fig. 8.2.2-1).

Cosmic rays associated with a flare or CME launched from near the central meridian of the Sun travel to Earth in a somewhat different manner. It is now understood that in this case most of the cosmic rays that reach Earth were accelerated in the shock wave associated with the CME after it has travelled to the vicinity of five solar radii from its source (Kahler 1994; Reames 1999), by which time it extends over a relatively large range of solar longitudes ($>120^\circ$). Cosmic rays are accelerated at many points in the shock over this wide range of solar longitudes, and a portion escapes into the heliospheric field ahead of the shock. They ultimately find their way to Earth via many different paths often involving significant diffusion, and the distance the cosmic rays travel before they reach Earth will vary from one particle to another. As a consequence the

intensity rises slowly, the particles are not well focused, and they are observed to arrive at the Earth in an approximately isotropic manner. At lower energies, particles continue to be accelerated in the outward moving shock wave (Sect. 8.2.2), which further extends the rise to maximum intensity (e.g. Fig. 8.2.2-2, 19 October, 1989). At higher energies, a Forbush decrease will often be associated with the passage of the shock wave past the Earth. The GLE of 25 July 1946 is a good example of this class of event (Fig. 8.2.1-1).

Figure 8.2.2-4 illustrates these propagation effects at low energies. The left hand graph gives the longitudes (on the Sun) of the solar activity that resulted in a large number of “gradual” events (plotted one event at a time). Clearly, gradual SEP events originate at longitudes that are distributed symmetrically about the central meridian of the Sun. The “impulsive events” are preferentially associated with acceleration on the far western portion of the solar disk, and extending to acceleration that has taken place behind the western limb. Similar distributions are observed for the high-energy GLE observed by neutron monitors.

Referring back to Fig. 8.3.1-2, note that the “hardest” spectra (high intensities near 1 GeV; not so intense at low energies) were those from events (1) and (13), both resulting from particle acceleration near the western limb of the Sun. The softer spectra events [e.g. (2), (3) and (4)] originated near the central meridian of the Sun (solar longitude near 0°). These associations between spectra and where they originate are evident for the 70 events in the GLE record.

8.3.3 *The Frequency of Occurrence, and the Detection of Historic SEP Events*

Figure 8.3.3-1 illustrates the rate of occurrence of SEP events since the sunspot minimum of 1954. The first 10 years are based on ionospheric and other measurements; the subsequent data are from satellite detectors. Note that this plots the >30 MeV fluence, i.e. this provides a semi-quantitative indication of the radiation dose and cosmogenic radionuclide production, from each SEP event in this interval. The graph shows that the production of cosmic rays in the vicinity of the Sun is a relatively common process and that it waxes and wanes throughout the 11-year solar cycle. Note also that the vertical scale is logarithmic, and that the SEP in this figure had fluences ranging over four orders of magnitude.

Figure 8.3.3-2 presents all the available information regarding the frequency of occurrence of SEP events, as a function of their >30 MeV omnidirectional fluence. These data are based on satellite, and nitrate paleo-cosmic ray data, as indicated. An upper limit has been obtained from the annual ^{10}Be data from Dye 3 in Greenland. A striking feature of this graph is the manner in which it steepens rapidly for >30 MeV fluences above 5×10^9 particles cm^{-2} . In other words, a SEP with a fluence of $>6 \times 10^{10}$ particles cm^{-2} is only likely to be seen once in 1,000 years. Satellite measurements have explained this effect (the Rosetta Stone principle

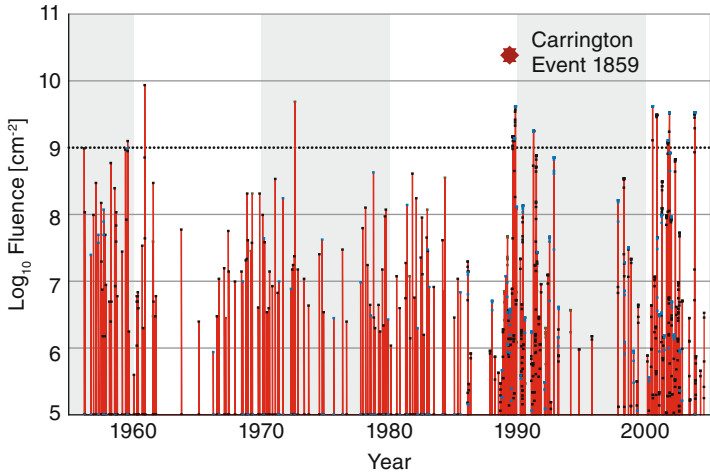
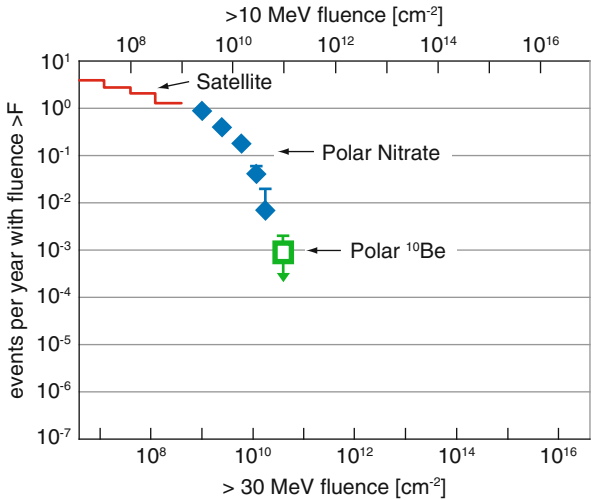


Fig. 8.3.3-1 The occurrence of SEP events since 1954 (Smart et al. 2006). The >30 MeV omnidirectional fluence of each event is given; the dotted line indicates the threshold for detection as an impulsive nitrate event; the star is the estimated >30 MeV omnidirectional fluence for the Carrington–Hodgson event of 1859 (See footnote, page 119)

Fig. 8.3.3-2 The cumulative probability of large SEP events derived from satellite and nitrate data (after McCracken et al. 2001) with an upper limit based on ^{10}Be data. Measurements of the cosmogenic radionuclides in moon rocks, and meteorites, are consistent with this distribution; in particular, they confirm the rapid increase in slope of the curve towards higher fluences



again) as a consequence of the dimensions of the acceleration region. This gives rise to a self-limiting process (the “streaming limit”) wherein the accelerated particles themselves modify the magnetic fields and inhibit the escape of the solar cosmic rays into the heliospheric field (Reames 1999). The nitrate data indicate that the Carrington–Hodgson flare of 1859 resulted in the greatest fluence in the vicinity of Earth (1.9×10^{10} particles cm^{-2}) since 1561 (See Footnote, page 119).

The physics of the streaming limit applies to single SEP events. There is abundant evidence that very large sunspot groups will repeatedly trigger major acceleration episodes, leading to a series of large SEP events within a single rotation of the Sun (27 days). For example, there was a series of three GLE within a week in November 1960. There was another series of three GLE in October 1989, and a series of four in May 1990. This indicates that there is always the possibility that there may be a series of SEP, each near the streaming limit, within a month or so. This effect is already included in the nitrate data in the distribution in Fig. 8.3.3-2 because individual events cannot be identified in these data. Nevertheless, we are dealing here with the statistics of small numbers, and the present day evidence makes it seem possible that a sunspot group may have developed in the past that was extremely prolific in its production of very large SEP events (See Footnote, page 119).

Over the 10,000 years of the Holocene, Fig. 8.3.3-2 suggests that an exceptional series of SEP events may have occurred once with a fluence of $>6 \times 10^{10}$ particles cm^{-2} . For a spectral shape like that of curve 3 in Fig. 8.3.1-2 we would expect it to have produced a 60% increase in the annual global averaged production of ^{10}Be (~200% for ^{36}Cl). If it had a spectrum like curve 1 the production would be up to a factor of ten higher; this seems highly unlikely, however, since the instrumental era has shown that the largest fluences are associated with solar events near the centre of the solar disk, which have spectra like curve 3 (Fig. 8.3.1-2). For ~4 year average ice samples, a 60% annual increase would result in a 15% enhancement in ^{10}Be which would probably be unrecognizable in the presence of the 11 year variation. For this reason, it appears that ^{36}Cl will give the best prospects for identifying extremely large SEP events in the past. Such large events would have serious practical implications (destruction of space assets and radiation hazards to space and high altitude flight) and it would be desirable to further refine Fig. 8.3.3-2 using the whole of the paleo-cosmic ray record from the past 10,000 years when it becomes possible.

References

- Dreschhoff GAM, Zeller EJ (1994) 415-year Greenland ice core record of solar proton events dated by volcanic eruptive episodes, Institute for Tertiary-Quaternary Studies-TER-QUA Symposium Series
- Ellison DC, Ramaty R (1985) Shock acceleration of electrons and ions in solar-flares. *Astrophys J* 298(1):400–408
- Forbush SE (1946) 3 unusual cosmic-ray increases possibly due to charged particles from the sun. *Phys Rev* 70(9–10):771–772
- IPCC (2007) Fourth Assessment on climate change – The Physical Science Basis Rep
- Kahler S (1994) Injection profiles of solar energetic particles as functions of coronal mass ejection heights. *Astrophys J* 428(2):837–842
- Lal D, Peters B (1962) Cosmic ray produced isotopes and their application to problems in geophysics. In: Wilson JG, Wouthuysen SA (eds) *Progress in elementary particle and cosmic ray physics*. North-Holland publishing company, Amsterdam

- Lovell JL, Duldig ML, Humble JE (1998) An extended analysis of the September 1989 cosmic ray ground level enhancement. *J Geophys Res-Space Phys* 103(A10):23733–23742
- McCracken KG, Dreschhoff GAM, Zeller EJ, Smart DF, Shea MA (2001) Solar cosmic ray events for the period 1561–1994 1. Identification in polar ice, 1561–1950. *J Geophys Res-Space Phys* 106(A10):21585–21598
- McCracken KG, Dreschhoff GAM, Smart DF, Shea MA (2004) A study of the frequency of occurrence of large-fluence solar proton events and the strength of the interplanetary magnetic field. *Solar Phys* 224(1):359–372
- McCracken KG, Moraal H, Stoker PH (2008) Investigation of the multiple-component structure of the 20 January 2005 cosmic ray ground level enhancement. *J Geophys Res – Part A – Space Phys* A12101:12118
- Meyer P, Parker EN, Simpson JA (1956) Solar cosmic rays of February, 1956 and their propagation through interplanetary space. *Phys Rev* 104(3):768–783
- Parker EN (1958) Dynamics of the interplanetary gas and magnetic fields. *Astrophys J* 128(3):664–676
- Parker EN (1963) *Interplanetary dynamical processes*. Wiley, New York
- Reames DV (1999) Particle acceleration at the Sun and in the heliosphere. *Space Sci Rev* 90(3–4):413–491
- Ryan JM, Lockwood JA, Debrunner H (2000) Solar energetic particles. *Space Sci Rev* 93(1–2):35–53
- Smart DF, Shea MA, McCracken KG (2006) The Carrington event: possible solar proton intensity-time profile. *Adv Space Res* 38(2):215–225
- Vitt FM, Jackman CH (1996) A comparison of sources of odd nitrogen production from 1974 through 1993 in the Earth's middle atmosphere as calculated using a two-dimensional model. *J Geophys Res-Atmos* 101(D3):6729–6739
- Webber WR, Higbie PR, McCracken KG (2007) Production of the cosmogenic isotopes ^3H , ^7Be , ^{10}Be , and ^{36}Cl in the earth's atmosphere by solar and galactic cosmic rays. *J Geophys Res – Part A – Space Phys* A10106:10107

Part III
Cosmogenic Radionuclides

Chapter 9

Introduction to Cosmogenic Radionuclides

In Part 2 we have discussed the cosmic radiation, one of the links between the Earth and the cosmos. Although called “cosmic radiation” it consists of particles originating mainly from our galaxy. These play a fundamental role in the generation of the cosmogenic radionuclides that this book is all about. The six following chapters (Chaps. 10–15) in this Part describe how and where cosmogenic radionuclides are produced, how nature deposited them in archives, and how we measure them.

The production of cosmogenic radionuclides requires that one element should be transformed into another; for example, cosmogenic ^{14}C is produced from ^{14}N . The alchemists of old days tried to do this sort of thing in the laboratory; their goal was to make gold from ordinary metals such as lead using chemical processes, and despite much effort they failed (Fig. 9-1).

Starting in the 1940s physicists learned how to build large particle accelerators that used strong magnetic and electric fields to accelerate protons, electrons, and heavier ions to very high energies. With them, they studied the composition and structure of the nuclei of atoms and how they are held together. They also discovered that the transmutation of one element into another was possible provided the energy of the accelerated particle was high enough. Other physicists were flying photographic emulsions on balloons to altitudes >30 km, where incoming cosmic rays were seen to collide with atoms in the emulsions, causing some of them to be transmuted into different elements. It became clear that high-energy particles, such as cosmic rays, could cause the conversion of the nucleus of one element into that of another. The dream of the alchemists had been realized but unfortunately it hasn't made the physicists rich; single atoms are produced at enormous cost. It was realized immediately that the incoming cosmic radiation had been producing “cosmogenic” nuclides for billions of years.

Generations of nuclear physicists have used their accelerators to carefully study all the processes involved in the interaction of high-energy particles with matter. To assist the design of new accelerators, and to conduct extremely complicated experiments, they developed mathematical models that simulate the interaction of fast particles with matter. Quickly, cosmic ray and radioisotope physicists used



Fig. 9-1 Alchemists working in a medieval laboratory on the conversion of base metals into gold (painted by Pieter Bruegel the Elder, 1558)

those mathematical programs to describe and quantify the production of the cosmogenic radionuclides in meteorites, in the atmosphere, and in the ground. The creation and use of these mathematical models were crucial steps in making the cosmogenic radionuclides into quantitative tools that would permit studying a large number of aspects of the environment and the past history of Earth.

In Chap. 10 we describe how the galactic cosmic ray particles interact with the atoms in the upper atmosphere, initiating a cascade of secondary protons, neutrons, and other particles. The use of Monte Carlo techniques to quantify the characteristics of the cascade is outlined, together with the subsequent nuclear processes that yield the production rates of cosmogenic radionuclides at every point in the atmosphere (as a function of solar activity and the strength of the geomagnetic dipole). In Chap. 11 we discuss the production of cosmic ray-induced radionuclides in the Earth (rocks, water, ice) as well as in extraterrestrial material (lunar rocks and meteorites). We finish the discussion of production mechanisms by outlining some which are not related to cosmic rays in Chap. 12.

In Chap. 13 we then ask the vital question – “what happens to the cosmogenic radionuclides after they are produced in the atmosphere?” How are they transported and distributed within the environmental system? This leads to the very important and practical questions of where we find them and how we measure them. As we will see, one of the unique properties of cosmogenic radionuclides is that they not

only provide us information about ongoing processes that augments information obtained directly by specialized instruments, but they also provide the means to study past environmental changes such as the long-term history of solar activity and its effects upon the climate of Earth. An important issue is therefore to find the natural cosmogenic archives where this information is stored and to learn how to decipher them (Chap. 14). Last but not least, Chap. 15 outlines the powerful analytical techniques that allow us to measure an extremely small number of cosmogenic atoms with high precision, without which all the information stored in the archives would be useless.

Chapter 10

Production of Cosmogenic Radionuclides in the Atmosphere

10.1 Introduction

In this chapter, we discuss what happens to a cosmic ray particle after it has successfully travelled through the heliosphere and the Earth's magnetic field and has reached the top of the atmosphere. The fact that it made it so far means that its energy is high (≥ 1 GeV). Depending on the degree of solar activity (Sect. 5.7, also Fig. 10.2.3-1), the majority of the low energy particles fail to get close enough to the Earth to interact with matter even in the polar regions.

Each nuclide is characterized by a specific number of protons and neutrons in its nucleus. To produce a cosmogenic nuclide in the atmosphere, it is necessary to change the number of nucleons in an atmospheric atom; for example, ^7Be , ^{10}Be , and ^{14}C are almost completely produced from atmospheric nitrogen (^{14}N) and oxygen (^{16}O) atoms. Since the binding energy for nucleons is large (>8 MeV), this never occurs as a result of thermal atmospheric collision processes. In the next two sections, we describe how a high-energy cosmic ray particle initiates a chain of nuclear interactions that ultimately result in the nuclei of atmospheric atoms being disrupted, yielding cosmogenic nuclei.

The nuclear processes initiated by the cosmic rays nearly always result in nuclei with atomic masses equal to or less than that of the target atmospheric nucleus. The only exception is neutron capture which increases the target mass by one as in the case of $^{35}\text{Cl}(n,\gamma)^{36}\text{Cl}$. Since the main constituents of the atmosphere are nitrogen and oxygen (Table 10.1-1 in Box 10.1.1), the majority of the cosmogenic radionuclides produced have masses below 16. Masses up to 40 are rare because the weight fraction of argon in air is only 1.3%. Heavier cosmogenic radionuclides are very rare because of the very small amounts of Krypton and Xenon in the atmosphere. Neglecting those with short half-lives leaves us with a very limited number of cosmogenic radionuclides produced in the atmosphere.

The interaction of a primary cosmic ray particle with the atmosphere consists of various physical processes leading ultimately to the dissipation of the particle's kinetic energy. All these processes have been studied extensively in the last several

Table 10.1-1 Some characteristics of the most abundant atmospheric components

Gas	Percentage by volume	Percentage by weight	Total weight (10^{20} g)
Nitrogen	78.09	75.51	38.65
Oxygen	20.95	23.15	11.84
Argon	0.93	1.28	0.655
Carbon dioxide	0.0386	0.059	0.030
Neon	0.0018	0.00125	0.000636
Krypton	0.0001	0.00029	0.000146
Xenon	0.000008	0.000036	0.000018

Source: <http://www.britannica.com/EBchecked/topic/108636/chemical-element/81245/The-atmosphere>

decades using accelerators and are generally well understood. It is this knowledge that is used to calculate the production rates of cosmogenic radionuclides. In Sect. 10.2, the interaction of cosmic rays with the atmosphere and the production of secondary particles (mainly neutrons and protons) are discussed. Many hundreds of interactions occur in each cascade, and we outline the complex physical models used to determine the manner in which the various secondary components develop and attenuate with depth. Then, in Sect. 10.3, we estimate the probability that a secondary particle interacting with oxygen, nitrogen, or argon produces a specific cosmogenic radionuclide. As we will see, the production rate depends on (1) the intensity of the incoming cosmic rays and their modulation by solar activity, and the intensity of the geomagnetic field and (2) the atmospheric depth and the geomagnetic latitude where the production takes place. The calculations show that the galactic cosmic rays are primarily responsible for the production of the cosmogenic nuclides, but that there can be a detectable contribution from solar cosmic rays during limited periods of time. Based on our present knowledge, they do not play a significant role when averaged over 10 years or more.

10.1.1 BOX The Earth's Atmosphere

The atmosphere consists of a mixture of gases that form a thin layer around the Earth and act as an interface between the Earth and space. Table 10.1-1 presents some characteristics of seven of the chemical elements present in the atmosphere.

It is interesting to note that the presence of ^{40}Ar is mainly the result of the radioactive decay of terrestrial ^{40}K , which was produced as part of the nucleosynthesis in a supernova explosion more than five billion years ago. Together with a cloud of interstellar gas and dust, it gave rise to our Sun and the planets. Its long half-life of $2.3 \cdot 10^9$ y means that it is still very common in the Earth's crust.

The main characteristics of the atmosphere are that the air gets thinner with increasing altitude and that there is no upper boundary. Figure 10.1-1 shows that both the density and the pressure decrease approximately exponentially with height. About half of the total amount of air is below 5.6 km.

On account of their different properties, the atmosphere is divided into five regions called the troposphere (lowest), stratosphere, the mesosphere, the

thermosphere, and the exosphere. As far as cosmogenic radionuclides are concerned, only the troposphere and the stratosphere matter. The tropopause, the boundary between the troposphere and the stratosphere, has a variable height, ranging from about 8–12 km above the poles to 16–18 km above the equator. The temperature decreases from sea level to the top of the troposphere by about 6.5°C per km (the lapse rate), which leads to thermal convection of the air. The stratosphere extends up to 50 km and the mesosphere to 85 km, with an increasing temperature resulting in vertical stratification.

The dependence of temperature, density, and pressure on the altitude above sea level, shown in Fig. 10.1-1, is called the standard atmosphere. In reality, all three parameters are subject to temporal and spatial variability caused by the local weather conditions. The upper atmosphere is strongly affected by solar activity.

The pressure is measured in several different units (g cm^{-2} , hectopascal or millibar, atmosphere) throughout this book. The following table provides the corresponding conversion factors. Hectopascal (hPa) and millibar are alternative names for the same pressure unit (Table 10.1-2).

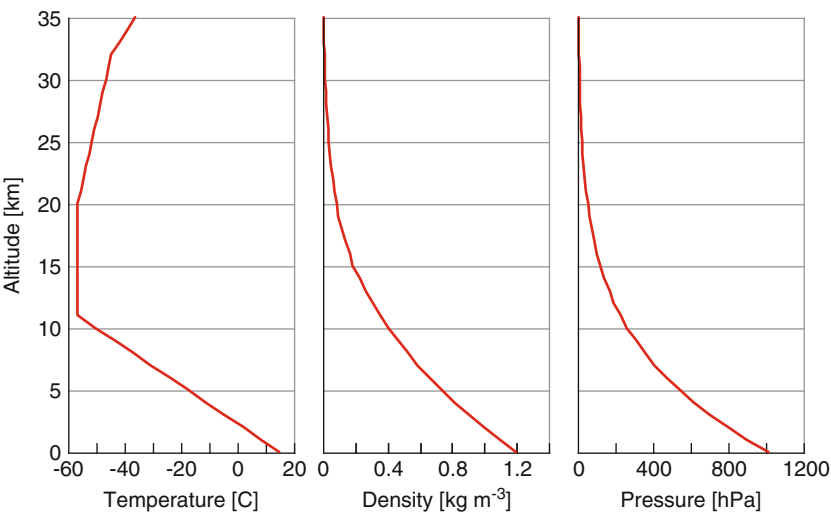


Fig. 10.1-1 Dependence of temperature, density, and pressure on the altitude in the standard atmosphere

Table 10.1-2 Conversion factors between some common units of atmospheric pressure

Units	g cm^{-2}	hPa or millibar	atm
g cm^{-2}	1	0.981	0.0009678
hPa or mb	1.0197	1	0.0009869
atm	1,033.2	1,013.25	1

10.2 Interaction of Primary Cosmic Rays with the Atmosphere

10.2.1 Production of Secondary Particles

Figure 10.2.1-1 illustrates a typical interaction of an incident primary particle with a nucleus of an atom in the upper atmosphere. A variety of secondary particles are produced which, in further interactions, produce additional particles and so on. A cascade develops until all the kinetic energy of the primary cosmic ray has been dissipated.

The kinetic energy of a cosmic ray nucleus is much greater than the nucleonic binding energy, and consequently primary alpha and heavier cosmic rays break up into their component nucleons during the first interaction. They then continue on their ways independently, each generating its own cascade. This permits an

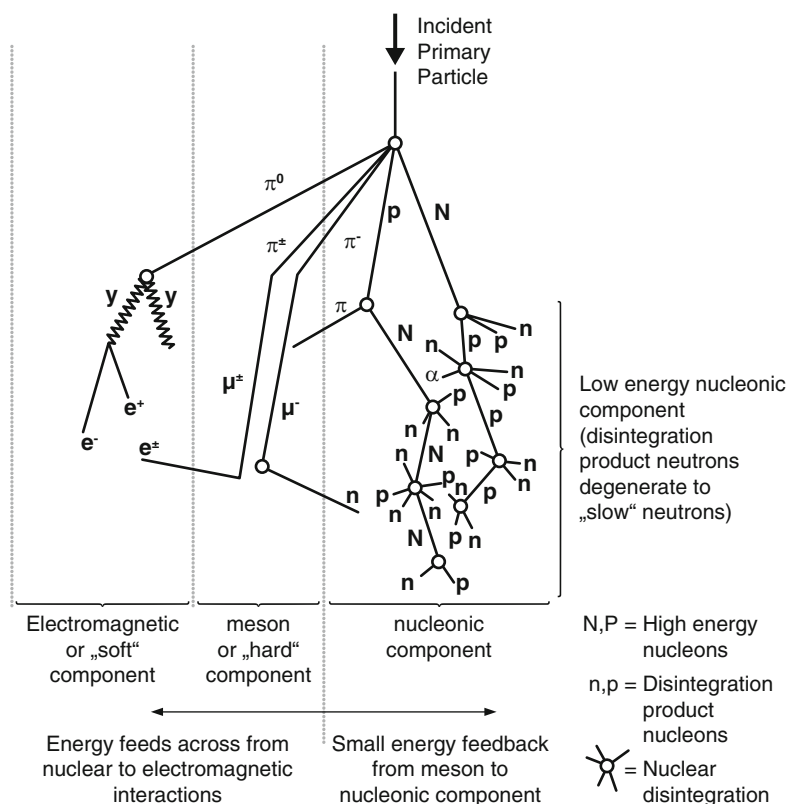


Fig. 10.2.1-1 Cascade of secondary particles produced by a primary cosmic ray particle incident on the top of the atmosphere. The cascade consists of a nucleonic, a mesonic and an electromagnetic component. The nucleonic component is the most important one for the production of cosmogenic radionuclides. It is also recorded by neutron monitors. After Simpson (2000)

important simplification of the analysis of the cosmogenic production process. Without loss of generality, we may compute the neutron and proton populations in the atmosphere, and the production of the several cosmogenic nuclides for a single incident proton. This we do in this and the next section. As discussed in Sect. 5.8.2, the shielding effect of the geomagnetic field is less for alpha and heavier cosmic ray particles than for a proton of the same energy per nucleon, and this has a significant effect on the production of the cosmogenic nuclides at low geomagnetic latitudes. We allow for this, and discuss some other practical matters in Sect. 10.4.

With regard to the secondary particles, we distinguish between a nucleonic component (n, p), a meson component ($\pi^0, \pi^\pm, \mu^\pm$), and an electromagnetic component ($e^\pm, \gamma$). The nucleonic component consists of high-energy protons and neutrons (indicated by P and N in Fig. 10.2.1-1) that are the result of nucleon–nucleon interactions, and low-energy nucleons (indicated by p and n) that originate from the disintegration of target nuclei hit by the cosmic ray particles and secondary nucleons. The particles of the meson component are characterized by very short half-lives and a lower probability (smaller cross-section) to interact with atoms. Most of the pions ($\pi^\pm$) decay into muons ($\mu^\pm$) with a half-life of 1.8×10^{-8} s, while the pi-nought mesons (π^0) disintegrate into γ -rays, contributing to the electromagnetic component. Although it seems that a high-energy μ -meson cannot travel more than 1 km in its mean lifetime of 2.2 μ s, many of them arrive at the Earth's surface some 20 km below the point where they were produced, a nice example of the theory of relativity that states that for high-energy particles travelling close to the speed of light with respect to our frame of reference, their clock appears to run slower, and therefore they live longer (see Box 5.3.1).

In the atmosphere, neutrons and protons dominate the production rate of cosmogenic radionuclides. However, due to their weaker interaction (larger mean free path), mesons play an important role deeper down in the Earth's crust.

In describing the development of the nuclear and meson cascades, we will use the interrelated concepts of nuclear cross-section and interaction mean free path. They are described in some detail in Boxes 10.2.1.1 and 10.2.1.2. Briefly, the cross-sections for the interaction of a hadron (strongly interacting particles, p , n , and K^- and π -mesons) with air determine the rate at which the cascade develops as it descends in the atmosphere. The cross-sections for multi-particle production specify the probabilities for the emission of the different particles produced in each interaction. Together, they provide the quantitative inputs that the mathematical models use to simulate the p , n , and pions produced by each cosmic ray. This is a probabilistic process; so the results will be different even for two identical cosmic rays. Nature and our mathematical models average over millions of individual cosmic rays to yield the results we observe.

10.2.1.1 BOX Cross-sections

Cross-sections are used to describe the probability that an interaction between a particle and a target will result in a specific process taking place. They provide the information that allows us to understand and model the interactions of cosmic ray particles with matter.

(continued)

Let us assume that a beam J of protons is crossing an infinitesimally thin layer of nitrogen gas with a thickness of dx (Fig. 10.2.1-2). Of the many reactions that occur when protons interact with nitrogen, let us consider a specific one, namely, the production of ^{10}Be . The concept of cross-section is that for each process, we attribute to each nitrogen atom a target area, or cross-section, σ . If the incoming proton hits this area, a ^{10}Be atom is produced; if it misses it, no ^{10}Be is produced. A cross-section, therefore, has the unit of an area. Since this area is very small, it is measured in barns ($1 \text{ b} = 10^{-24} \text{ cm}^2$). Carefully designed experiments have allowed the cross-sections of a great number of nuclear reactions to be measured and used in calculations such as those described here.

We now compute the probability, dq , that a proton crossing the nitrogen layer of area F produces a ^{10}Be atom. If the density of atoms in the volume Fdx is n , then the total cross-section available to produce a ^{10}Be atom is $nF\sigma dx$.

The probability dq that the specified reaction will take place is, therefore, the ratio of the sum of the individual cross-sections, $nF\sigma dx$, and the area of the slab, F :

$$dq = nF\sigma dx / F = n\sigma dx \quad (\text{B10.2.1-1})$$

Note that it is important that the layer dx is thin; otherwise the cross-section areas of individual atoms could overlap and thus reduce the sum of the individual cross-sections.

The attenuation of the proton beam intensity due to ^{10}Be production in distance dx is then

$$dJ = -Jdq = -Jn\sigma dx, \quad (\text{B10.2.1-2})$$

which is rewritten as the differential equation

$$dJ/dx = -Jn\sigma = -J/\Lambda_a, \quad (\text{B10.2.1-3})$$

where we define the mean free path by $\Lambda_a = 1/(n\sigma)$. In our example, Λ_a is the mean distance a proton travels before it produces a ^{10}Be atom (see Box 10.2.1.2).

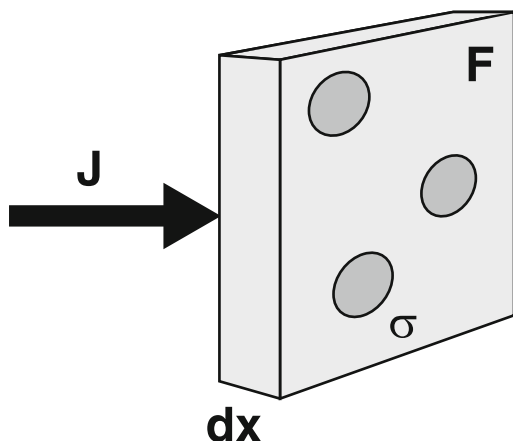
From the definition, it is clear that Λ_a is large if the cross-section is small; that is, the proton has to travel a long way before the specified interaction occurs.

Provided Λ_a is independent of x , the solution of (B10.2.1-3) is

$$J(x) = J_0 e^{-x/\Lambda_a} \quad (\text{B10.2.1-4})$$

It is sometimes useful to eliminate the atomic density n (number of atoms per volume) from the equation for the mean free path. The mass of each target

Fig. 10.2.1-2 Flux J of protons impinging on an infinitesimally thin layer dx of air. If the proton hits one of the cross-sections attributed to each atom, a specific reaction (e.g., production of ^{10}Be atom) takes place



atom (nitrogen in the example) is A/L , where L is the Loschmidt or Avogadro number and A is the atomic weight. If the mass density of the material in the box in Fig. 10.2.1-2 is ρ , then the mass of the target in the volume can be written in two ways:

Mass = $Fdx \, nA/L = \rho Fdx$ and simplifying, $n = L\rho/A$, and from this, the mean free path is $\Lambda_a = 1/(n\sigma) = A/L\rho\sigma$. Substituting this for Λ_a in (B10.2.1-4) gives

$$J(x) = J_0 e^{-L\sigma\rho x/A} \quad (\text{B10.2.1-5})$$

Equation (B10.2.1-5) shows that J decreases as function of ρx , that is, by the amount of matter in g cm^{-2} which the cosmic ray particles pass through. In many applications, the density of the absorbing material is constant, and (B10.2.1-5) shows that J decreases as a simple exponential function of distance into the absorber. The situation is very different in the atmosphere: as described in Box 10.1.1, the density of the atmosphere varies greatly between high altitudes and ground level. Thus, $J(x)$ decreases increasingly rapidly with depth because both x and density ρ increase monotonically from the top of the atmosphere to the bottom.

It is, therefore, desirable to define a new variable $X = \int \rho dx$, the “atmospheric depth”, defined as the amount of mass above any point in the atmosphere (measured in g cm^{-2}). This approach is similar to the concept of atmospheric pressure. Dividing both sides of (B10.2.1-3) by ρ , substituting X for ρx , and defining a new mean free path

$$\Lambda = \rho/(n\sigma) = A/L\sigma, \quad (\text{B10.2.1-6})$$

the differential equation (B10.2.1-3) becomes

(continued)

$$dJ/dX = -J/\Lambda \quad (\text{B10.2.1-7})$$

Solving this equation, we get

$$J(X) = J_0 e^{-X/\Lambda}, \quad (\text{B10.2.1-8})$$

where J decreases as a simple exponential function of the atmospheric depth (usually in g cm^{-2}). This equation, and the mean free path defined by (B10.2.1-6) will be used throughout this book.

$J(X)$ in (B10.2.1-8) is the intensity of the proton beam after reaching an atmospheric depth of X . Differentiating gives the number of collisions per unit change in atmospheric depth, which in this example is the production rate, P , of ^{10}Be atoms per gram cm^{-2} and second. Thus,

$$P(x) = dJ/dx = (L\sigma/a)J(x) \quad (\text{B10.2.1-9})$$

For particle energies much larger than the typical binding energy of nucleons within a nucleus (8 MeV), the collision of a hadron with the nucleus can be treated as an interaction between the hadron and the individual nucleons. Consequently, the hadron–nucleus interaction is basically a function of the nucleon–nucleon cross-section. This is illustrated in Fig. 10.2.1-3, which shows the proton–proton and the proton–air cross-section. After a steep rise at 1 GeV (not shown), it stays almost constant over several orders of magnitude. A detailed analysis of such an interaction reveals the following main results:

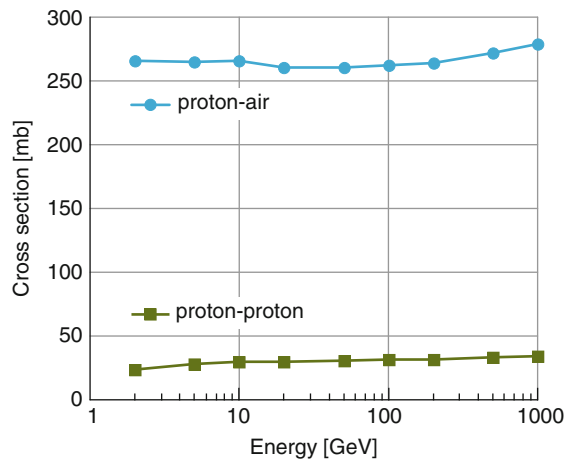


Fig. 10.2.1-3 Proton–proton and proton–air cross-sections. Both cross-sections show almost no energy dependence above 1 GeV. The cross-sections are measured in millibarns (10^{-27} cm^2)

Most secondary particles are emitted in the direction of motion of the incident nucleon, and little momentum is transferred perpendicular to it. This means that:

1. The mean free paths of the high-energy proton and neutron cascades are relatively independent of atmospheric depth.
2. The energy distribution of the high-energy tail of the primary proton flux follows a power law with $\gamma \approx 2.7$. We will see later that the shape of the energy distribution is independent of the atmospheric depth X . The only change with increasing depth is an exponential decrease in the total flux. For $\Lambda = 160 \text{ g cm}^{-2}$, the flux at sea level is about 640 times smaller than that at the top of the atmosphere ($e^{1033/160}$).
3. The broadening of the cascade is mainly due to the mesons which transfer energy and momentum perpendicular to the direction of incidence (Fig. 10.2.1-1). Their weaker interaction with matter is reflected in a larger free mean path ($\sim 250 \text{ g cm}^{-2}$).

It is important to note that:

1. The production rate at atmospheric depth X (per gram of added depth) is determined by the intensity of the particle flux $J(X)$ at that depth, times the cross-section σ , times the number of atoms in a gram of the target material L/A :

$$P(x) = dJ/dx = (L\sigma/a)J(x) \quad (\text{B10.2.1-9})$$

(see Box 10.2.1.1)

2. The cross-sections for a specific reaction are energy dependent.
3. The intensity $J(X)$ and production rate $P(X)$ are simple exponential functions of the atmospheric depth X . The distance is not important: it is the number of target atoms that counts. This means, for example, that the number of nuclear reactions between altitudes of 20 and 10 km is about the same as that between 6 and 3 km for the same cosmic ray intensity.

$$J(X) = J_0 e^{-X/\Lambda} \text{ (see Box 10.2.1.1)}$$

10.2.1.2 BOX Mean Free Path

In units of atmospheric depth, the mean free path Λ is defined as

$$\Lambda = 1/(np\sigma) = A/L\sigma \text{ (see (B10.2.1-6) in Box 10.2.1.1).}$$

Here, we show that Λ is indeed the mean free path a particle travels before undergoing a specific reaction described by the cross-section σ . Using (B10.2.1-8)

$$J(X) = J_0 e^{-X/\Lambda},$$

the probability dp that the particle interacts between X and $X + dX$ is then given by

(continued)

$$dp(X) = [J(X) - J(X + dX)]/J_0 = (1/\Lambda)e^{-X/\Lambda}dX \quad (\text{B10.2.1-10})$$

The average atmospheric depth travelled by a proton between collisions is, therefore,

$$\langle X \rangle = \int_0^\infty X dp(X) = \int_0^\infty \frac{X}{\Lambda} e^{-X/\Lambda} dX \quad (\text{B10.2.1-11})$$

For protons in the atmosphere, the mean free path Λ can be calculated using (B10.2.1-6): With $A = 29$ g for air, $L = 6 \times 10^{23}$, and $\sigma = 265$ mb (Fig. 10.2.1-3), we get $\Lambda = 182$ g cm⁻², which is close to the observed 160 g cm⁻² and means that on average the proton cascade involves about seven interactions before reaching the Earth's surface. In rock, 182 g cm⁻² corresponds to about 60 cm depending on the density.

10.2.2 Ionization and Excitation Processes

Charged particles passing through matter lose energy due to Coulomb interactions between the particle and the electrons of the neighbouring atoms. For low-energy particles, ionization is the main process that contributes to energy loss. The mean energy loss (stopping power) is described by the Bethe–Bloch formula.

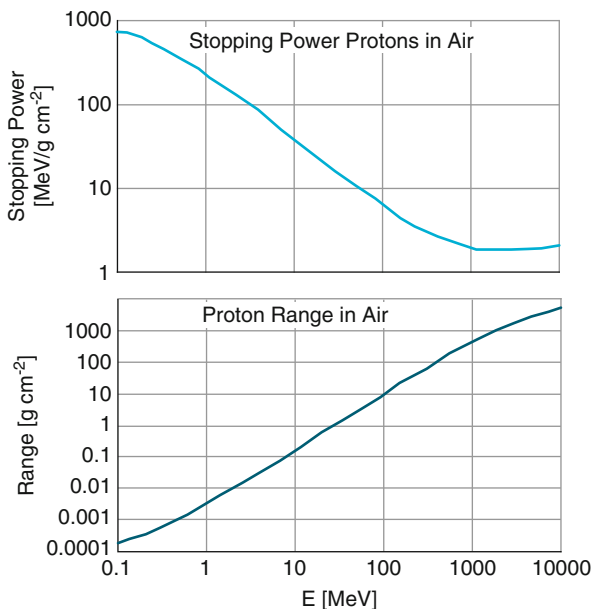
$$\frac{dE}{dx} = -\frac{4\pi}{m_e c^2} \frac{nz^2}{\beta^2} \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \left[\ln \left(\frac{2m_e c^2 \beta^2}{I(1 - \beta^2)} \right) - \beta^2 \right], \quad (10.2.2-1)$$

where

β	$=v/c$
v	Velocity of the particle
E	Energy of the particle
x	Distance travelled by the particle
c	Speed of light
ze	Particle charge
e	Charge of the electron
m_e	Rest mass of the electron
n	Electron density of the target
I	Mean excitation potential of the target
ϵ_0	Permittivity of free space

The second term in (10.2.2-1) shows that the energy loss due to ionization depends mainly on the electron density of the medium n , and the squares of the velocity ($1/\beta^2$) and charge (z^2). The fourth term describes the effects that occur at relativistic speeds (see below). The other terms are constants.

Fig. 10.2.2-1 Stopping power (energy loss per g cm^{-2}) and range (g cm^{-2}) of protons in air. The stopping power decreases with increasing proton energy and reaches a broad minimum of about $2 \text{ MeV g}^{-1} \text{ cm}^2$ at 1 GeV. Such particles are called “minimum ionizing”. The range of a proton increases monotonically with energy



In the non-relativistic case, the Bethe–Bloch formula can be rewritten showing that the energy loss dE/dx depends on the mass m of the particle, the square of its electrical charge z , and the reciprocal of the kinetic energy E :

$$\frac{dE}{dx} \propto \frac{mz^2}{E} \quad (10.2.2-2)$$

At relativistic speeds, the fourth term in the Bethe–Bloch formula shows that the energy loss increases slowly with increasing velocity. Figure 10.2.2-1 gives the stopping power and the ranges of protons in air, and this illustrates how the effects combine. At first, the stopping power decreases as the reciprocal of the square of the velocity, it reaches a minimum, and then rises again logarithmically. The range increases approximately exponentially with increasing energy. Due to the large stopping power at low energies, the range of a 1 MeV proton is very short (0.003 g cm^{-2}).

Examination of Fig. 10.2.2-1 provides the following insight into the effects of ionization loss on the incoming cosmic rays, and the nucleonic cascade.

1. The lower energy cosmic rays ($E < 100 \text{ MeV/nucleon}$) lose their energy very quickly ($>10 \text{ MeV g}^{-1} \text{ cm}^2$) and rapidly come to the end of their range high in the stratosphere ($X < 10 \text{ g cm}^{-2}$). By way of comparison, $\sim 2 \text{ GeV/nucleon}$ cosmic rays are “minimum ionizing”, and will only lose $<10\%$ of their energy by ionization before initiating a nucleonic cascade (which occurs before they reach an atmospheric depth of $\sim 100 \text{ g cm}^{-2}$).

2. Low-energy protons are less effective than neutrons in the production of cosmogenic nuclides because (1) in air, they lose energy rapidly through ionization and (2) in order to induce nuclear reactions, they must first penetrate the Coulomb barrier of the target nucleus.

Having discussed some of the basic physical processes that determine the interaction of cosmic ray particles with matter in general and the atmosphere in particular, we are now ready to discuss the production of the cosmogenic radionuclides themselves. The first fundamental step is to identify the particles responsible for the production of cosmogenic radionuclides and to determine their differential energy fluxes within the atmosphere. As outlined in the introduction (Sect. 10.1), this is a difficult task because every high-energy primary particle starts a cascade of secondary particles, which in turn generate a third generation of particles and so on. In this way, a complex cascade develops all the way through the atmosphere until the energy is dissipated (Fig. 10.2.1-1). In the next section, we describe how a “Monte Carlo” code is used to simulate this process, and present the proton and neutron fluxes it produces.

The second step (Sect. 10.3) is then to use the simulated fluxes of protons and neutrons at any point in the atmosphere to calculate the production of each nuclide of interest. To this end, the simulated fluxes are multiplied by the appropriate cross-sections and integrated over all energies.

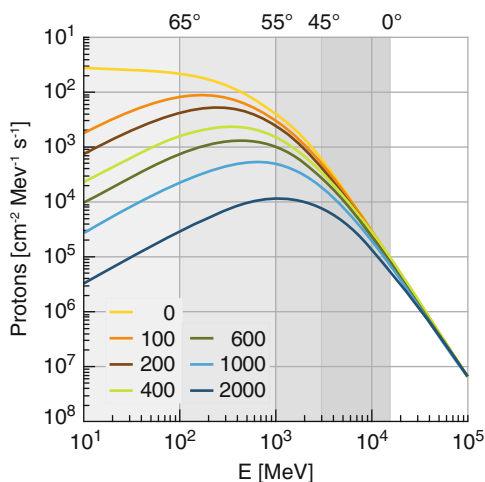
10.2.3 Simulated Atmospheric Proton and Neutron Fluxes

The development of nucleonic and electromagnetic cascades in matter (e.g., the atmosphere) is a statistical process. The result of each particle interaction is not known precisely; all we know is the probability that various particles will be produced. What we see in nature is the average over millions of individual statistical encounters. Statistical processes are common in nuclear physics, and the so-called Monte Carlo calculation methods have been developed to simulate the statistical games that the nuclear particles play.

A number of Monte Carlo codes (e.g., GEANT, MCNP, and Fluka) have been developed to assist in the design of detector systems in high-energy physics experiments. Over the years, they have been improved and extended by worldwide collaborations of many scientists and are now applicable to many more fields such as space and cosmic ray physics, shielding, detector development, medical physics, and radio-biology.

In the following, we give a short description of the use of one of them, GEometry ANd Tracking (GEANT; <http://www.geant4.org/geant4/index.shtml>). GEANT was developed at CERN to simulate the passage of particles through matter. Given the energy and direction of motion of a primary cosmic ray proton, GEANT tracks its penetration into the atmosphere, simulating the nuclear interactions and the generation of secondary particles (p , n , α , e , π , μ). Each new secondary particle is tracked individually and its interactions (ionization, energy loss, elastic and inelastic

Fig. 10.2.3-1 Differential energy spectra for primary galactic cosmic ray protons after passing through the heliosphere. The seven curves represent different levels of solar activity expressed by the solar modulation function Φ given in the inset in MeV. The geomagnetic latitude scale along the top indicates the lowest proton energy that can reach the top of the atmosphere from the vertical at that latitude (see Fig. 5.8.2-2)



nuclear reactions, and decay) are simulated until all the energy is dissipated. GEANT contains all the necessary physics to do this; in particular, it simulates both the electromagnetic and hadronic interactions in matter. Parameters which cannot be calculated (e.g., cross-sections) are tabulated in extensive libraries. Overall, the program includes more than 200,000 lines of code (Brun et al. 1987).

The GEANT code was used to produce the proton and neutron flux spectra needed to calculate the production rates of cosmogenic nuclides (Masarik and Beer 1999, 2009) and those used in the following discussions. The atmosphere was approximated by 34 concentric shells, each having a thickness of 30 g cm^{-2} (small compared to the nucleon–nucleon mean free path). Bearing in mind the geomagnetic cut-off effects (Sect. 5.8), each hemisphere of the world was divided into nine bands, each 10° wide in geomagnetic latitude. Thus, all the points on the Earth in each band had approximately the same geomagnetic cut-off rigidities (in absolute value, and as a function of arrival direction), and all saw the cosmic ray spectrum integrated from those rigidities to infinity. For each of these bands, 5,000,000 simulations were carried out selecting primary protons and alpha particles to represent each specific differential spectrum as outlined below, and selecting the arrival direction at the top of the atmosphere at random within the upward facing hemisphere (i.e., down to zenith angles of 90°).

The primary differential cosmic ray spectra used in the simulations are illustrated in Fig. 10.2.3-1. The top curve is the Castagnoli and Lal (1980) estimate of the “Local Interstellar Spectrum” (LIS), as discussed in Sect. 5.7.2. The six spectra below show the effect of the solar modulation (Sect. 5.7) on the proton flux for different values of the modulation function Φ . Note that the differences in the spectra between $\Phi = 0 \text{ MeV}$ and $\Phi = 2,000 \text{ MeV}$ are quite large, and we will see that this causes order of magnitude and greater differences in the fluxes of protons and neutrons in the atmosphere.

As discussed in Sect. 5.7, there are no direct measurements of the LIS, and as a consequence, there are a number of estimates as shown in Fig. 5.7.2-1. It is also likely that there will be improvements in the estimates of the LIS in the future, particularly if a new “Grand Minimum” of solar activity should occur. Each formulation of the LIS implies different production rates for neutrons, protons, and cosmogenic nuclides. As discussed in Sect. 5.7, the choice of LIS affects the estimates of the modulation function made from the observed data. Where necessary, conversion algorithms can be developed that allow the consequences of different LIS to be explored (see Fig. 5.7.2-1) (Usoskin et al. 2005; Steinhilber et al. 2008).

The effect of the geomagnetic field on the primary cosmic ray spectrum is quite different from that of solar modulation. As discussed in Sect. 5.8, cosmic rays with rigidities below the “cut-off rigidity” are prevented completely from reaching the top of the atmosphere. Equation (5.8.2-1) shows that the cut-off rigidity depends strongly upon latitude λ and on M , the magnetic dipole moment of the Earth. The corresponding cut-off energy is given by (5.3-7B). Since the cosmic ray flux is the integral of the spectrum (Fig. 10.2.3-1) from the cut-off energy to infinity, it is clear that it will vary strongly in response to changes in M . Likewise, it will depend strongly on the value of the cut-off rigidity: the higher the cut-off, the lower the flux. As a consequence, in the following, we will be studying the dependence upon λ and M of both the atmospheric nucleonic populations and the cosmogenic production rates.

Figure 10.2.3-1 depicts the situation for vertical incidence only. While protons need an energy of at least 13.9 GeV to reach the equator, only 3 GeV is required at 45°, and at 55°, 1 GeV is sufficient. The figure clearly shows that as a result of the different cut-off energies, the effect of solar modulation is largest at high geomagnetic latitudes and becomes comparatively small at the top of the atmosphere in the equatorial regions.

We note in passing that having selected the arrival direction, all protons below the cut-off rigidity applicable to that direction were rejected from the simulation. Equation (5.8.2-1) shows that near the equator, the cut-off rigidities range from ~ 60 GV from the eastern horizon ($\vartheta = 0^\circ$) to ~10 GV from the west ($\vartheta = 180^\circ$). Figure 10.2.3-1 shows that this implies a cosmic ray flux that differs by a factor of ~50 from the east to the west. Clearly, this results in a substantial asymmetry in the nucleonic cascades (and the production of the cosmogenic nuclides) from east to west near the equator. Equation (5.8.2-1) shows that this “east–west” asymmetry disappears rapidly towards the polar regions so that above 45° latitude, the proton fluxes are essentially the same over the whole upper hemisphere.

Over time scales of a century or more, the strength of the geomagnetic field varies, and this, in turn, affects the cut-off rigidities (Sect. 5.8). For the purposes of these simulations, the dipole approximation that was used is parameterized in terms of the geomagnetic moment M in units relative to the present value. As we can see from Eq. 5.8.2-2, the vertical Stoermer cut-off at the equator varies from 0 to 29.8 GV for $M = 0$, and 2, and examination of Fig. 10.2.3-1 shows that this

would have a very great effect on the flux of cosmic rays reaching the equatorial regions, and consequently on the proton and neutron populations in the atmosphere.

In the following, we use some of the GEANT simulations to demonstrate the nature of the particle fluxes that are responsible for the production of the cosmogenic nuclides (Sect. 10.3).

In Fig. 10.2.3-2, the simulated proton spectra are shown for the atmospheric depths 30, 60, 90, 120, 300, 690, and 1,020 g cm^{-2} . The depth of 300 g cm^{-2} corresponds to typical aeroplane cruising heights (30,000 feet or 10,000 m), and 690 g cm^{-2} approximates the altitude of the Climax neutron monitor. At a first glance, it seems that all spectra have the same shape, a slight increase for energies below 200 MeV followed by rather steep decrease. Closer inspection reveals that the difference in the fluxes for low energies is about a factor 2,000 between the top of the atmosphere (30 g cm^{-2}) and sea level (1,020 g cm^{-2}), while it decreases by a factor of $>20,000$ at 10^3 MeV. The reason for this effect is that secondary protons with relatively low energies can be produced for a number of generations of the cascade from interactions taking place higher up in the atmosphere, while high-energy protons occur only during the initial stage of the cascade and therefore become rather rare with increasing depth.

Another interesting feature visible at the depths of 30 and 60 g cm^{-2} is the presence of the clear bumps starting at 100 and 200 MeV respectively. Figure 10.2.2-1 shows that the range of 100 and 200 MeV protons is $\sim 10 \text{ g cm}^{-2}$ and $\sim 30 \text{ g cm}^{-2}$ respectively, indicating that the bump in the 30 g cm^{-2} curve is due to primary cosmic ray protons that can reach this depth.

Figure 10.2.3-3 now adds the neutron fluxes to the proton fluxes. All the spectra are simulated for the geomagnetic latitude band of $80\text{--}90^\circ$; there is no solar modulation and the present magnetic field intensity is assumed. The proton fluxes (solid lines) show the same features as already discussed in Fig. 10.2.3-2. Note, however, that below 500 MeV, the neutron fluxes (dotted lines) are considerably

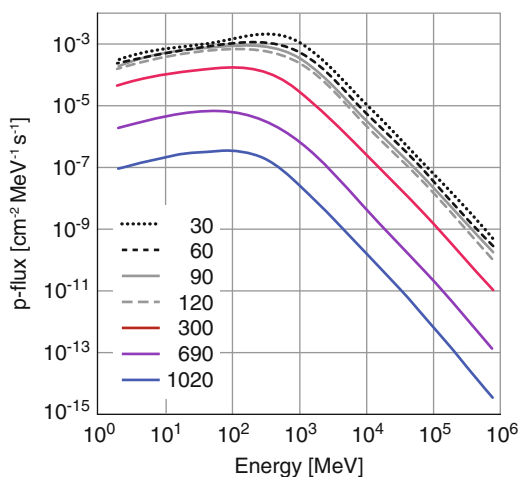
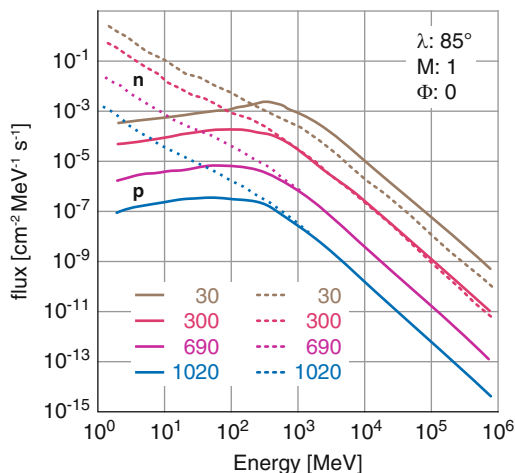


Fig. 10.2.3-2 Simulated proton fluxes for different atmospheric depths ranging from 30 to 1,020 g cm^{-2} for the present magnetic field ($M = 1$), no solar modulation ($\Phi = 0 \text{ MeV}$), and latitude 85°

Fig. 10.2.3-3 Simulated differential energy spectra for protons (*solid lines*) and neutrons (*dotted lines*) for different atmospheric depths ranging from 30 to $1,020 \text{ g cm}^{-2}$ for the present magnetic field ($M = 1$), no solar modulation ($\Phi = 0 \text{ MeV}$), and geomagnetic latitude 85°



larger than those of the protons (more than a factor of 100 at 10 MeV). For energies above 1 GeV, the situation changes with depth. For example, at the top of the atmosphere (30 g cm^{-2}), protons are more abundant due to the primary proton flux and the fact that the cascade is not yet fully developed. Below 300 g cm^{-2} , the neutron flux approximates the proton flux.

Figure 10.2.3-3 provides some important insights into the production of the cosmogenic nuclides in the atmosphere (Sect. 10.3). The fluxes of neutrons and protons are much higher in the stratosphere ($<400 \text{ g cm}^{-2}$); thus, this is where the production rates are the highest. The neutron fluxes are higher than proton fluxes for $E < 500 \text{ MeV}$ because the high rates of ionization loss (Fig. 10.2.2-1) quickly remove the low-energy protons, while the neutrons are unaffected. As a consequence, neutrons are more important for the production of the cosmogenic nuclides at these altitudes.

Figure 10.2.3-4 plots the dependence of neutron flux upon geomagnetic latitude. The fluxes are simulated at the depth of 300 g cm^{-2} (typical cruising altitude of airplanes) for a solar modulation function $\Phi = 550 \text{ MeV}$ (typical of the present) and the present geomagnetic field intensity ($M = 1$). While the flux is independent of latitude above 20 GeV, the curves start to deviate gradually more and more from one another towards the poles. For example, if we take the latitudinal band $60^\circ\text{--}70^\circ$ as a reference, the flux in the $0^\circ\text{--}10^\circ$ band is the first to deviate (highest cut-off rigidity), followed by the $10^\circ\text{--}20^\circ$ band, and so on. That is, the dependence of the neutron flux upon latitude increases strongly as a function of decreasing energy. This will be seen to result in a strong latitude dependence of the cosmogenic production rate.

Figure 10.2.3-5 investigates the effect of solar modulation on the proton and neutron fluxes. We display the highest latitude band ($80^\circ\text{--}90^\circ$) at the top of the atmosphere ($0\text{--}30 \text{ g cm}^{-2}$) for the present geomagnetic field intensity ($M = 1$). The protons (solid lines) and neutrons (dotted lines) are plotted for $\Phi = 0, 500$, and $1,000 \text{ MeV}$. The general behaviour of p and n is the same as we discussed in the previous figure. As expected from the primary proton spectrum (Fig. 10.2.3-1), the

Fig. 10.2.3-4 Differential energy spectra of neutrons for geomagnetic latitudes from 5° to 85° for the present magnetic field ($M = 1$), solar modulation ($\Phi = 550$ MeV), and atmospheric depth 300 g cm^{-2} . Note that the curves for latitudes $>65^\circ$ overlap

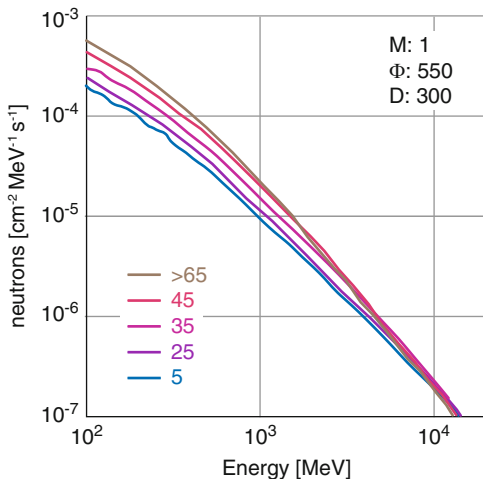
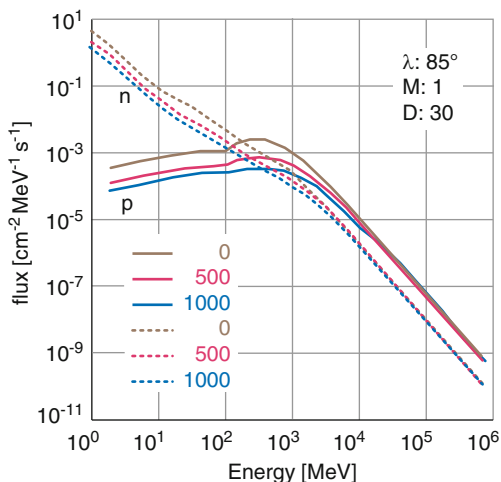


Fig. 10.2.3-5 Differential energy spectra of proton (solid lines) and neutron (dotted lines) fluxes for different solar modulation ($\Phi = 0, 500$, and $1,000$ MeV) for the present magnetic field ($M = 1$), geomagnetic latitude 80° – 90° , and atmospheric depth 0 – 30 g cm^{-2}



effect of the solar modulation is largest for low energies and has essentially disappeared at 10 GeV (10^4 MeV).

Finally, we explore the effect of changes in the geomagnetic field intensity upon the proton and neutron fluxes. For this purpose, we choose again an atmospheric depth of 300 g cm^{-2} and a solar modulation potential $\Phi = 550 \text{ MeV}$ for the latitude band 0° – 10° , where the effect will be largest. The result of this simulation is shown in Fig. 10.2.3-6a. Again the effects are marginal for energies above about 15 GeV and the neutrons are dominant for energies below 1 GeV . It is clear that changing the geomagnetic intensity from 0 to 0.5 has a much larger effect than doubling the present intensity from 1 to 2 . Since the transition region between 80 and 8 GeV shows some interesting features related to the dependence of the

geomagnetic cut-off on the angle of incidence, an enlargement of the area indicated by the frame is discussed in some more detail in Box 10.2.3.1.

10.2.3.1 BOX Geomagnetic Modulation of Atmospheric Neutron and Proton Fluxes

Figure 10.2.3-6b shows a blow up of Fig. 10.2.3-6a in the energy range where the geomagnetic cut-off takes effect. The figure shows that above 60 GeV, the proton and neutron fluxes are independent of the geomagnetic field strength. The top solid line shows the proton flux for no field ($M = 0$). From Sect. 5.8.2 and Fig. 5.8.2-2, we know that the cut-off energy for protons entering the atmosphere from the vertical (for $M = 1$) is about 14 GeV, and that it is up to 60 GeV for particles coming from the east and ~10 GeV for those coming from the west. In Fig. 10.2.3-6b, this is reflected in the fact that the proton flux for $M = 1$ starts to deviate from the $M = 0$ line at about 25 GeV and becomes almost parallel with the $M = 0$ curve at about 10 GeV. It is approximately half way between these two curves at ~15 GeV. For $M = 0.5$ and $M = 2$, we observe a very similar behaviour at half respectively and double the energy. The effect of the different geomagnetic intensities on the neutron flux starts at the same energies but occurs in a much smoother manner.

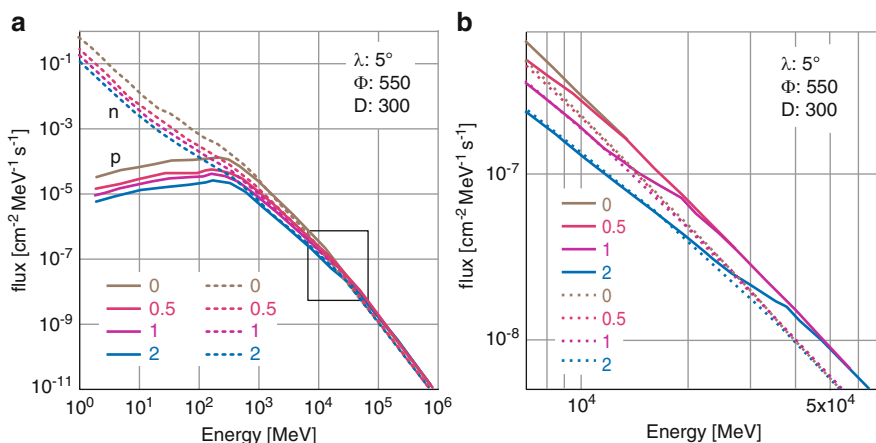


Fig. 10.2.3-6 (a) Differential energy spectra of protons (solid lines) and neutrons (dotted lines) for different geomagnetic dipole moments (0–2 times the present value) for solar modulation $\Phi = 550$ MeV, geomagnetic latitude 5° , and atmospheric depth 300 g cm^{-2} . (b) Enlargement of (a) for the energy range 800 MeV–70 GeV

10.3 Production of Cosmogenic Radionuclides in the Atmosphere

As will be discussed in Sect. 10.3.3, the cosmogenic nuclides are produced mainly by the interaction of neutrons and protons with nitrogen, oxygen, and argon, the main constituents of the atmosphere. The galactic cosmic rays entering the atmosphere do not normally contain neutrons, except for neutrons present in the nuclei of helium and heavier elements. On very rare occasions, there is a short-lived pulse of medium-energy neutrons associated with a solar energetic particle event (Sect. 8.2.2); however, the majority of those neutrons decay before reaching the Earth (neutron half-life = 10 min). As discussed in the previous section, an incoming cosmic ray initiates a nucleonic cascade in the atmosphere and a large variety of secondary particles is produced. Charged particles such as protons are slowed down by Coulomb interactions (Fig. 10.2.2-1). Since this effect is particularly strong at low energies, the shape of the proton differential energy spectrum is strongly modified at low energies (Fig. 10.2.3-3).

Figure 10.3-1 displays the integral proton flux with energies >20 MeV for the present-day geomagnetic field intensity and an estimated long-term average solar modulation function Φ of 550 MeV, as a function of geomagnetic latitude and atmospheric depth expressed in g cm^{-2} . Note that the flux and the depth scales are logarithmic. At the top of the atmosphere, the proton flux starts to decrease slowly

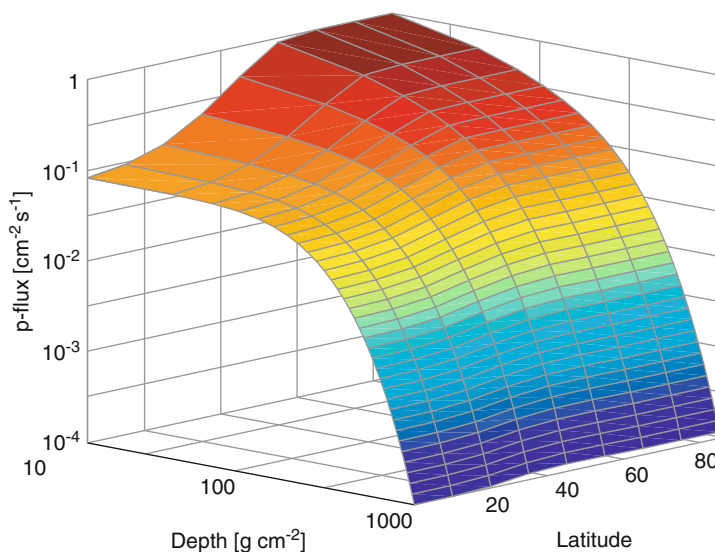


Fig. 10.3-1 Integral fluxes of protons with energies above 20 MeV for the present geomagnetic field ($M = 1$) and $\Phi = 550$ MeV. At the top of the atmosphere, the proton flux reaches values between 0.1 and 1 proton $\text{cm}^{-2} \text{s}^{-1}$ depending on the geomagnetic latitude and decreases rapidly with increasing depth by about three orders of magnitude

With increasing depth, the slope gets continuously steeper. The effect of the geomagnetic shielding is strongest at the top of the atmosphere and leads to a decrease in the proton flux by a factor 12 when moving from high to low geomagnetic latitude. This latitudinal effect decreases with increasing atmospheric depth to a factor of 1.2 at sea level.

The neutron flux exhibits a rather similar behaviour with latitude and depth, and Fig. 10.3-2 presents the neutron/proton ratio to illustrate some subtle but important points. At the top of the atmosphere, this ratio is 1.6 at low geomagnetic latitudes and 1 at high latitudes. Approaching sea level, the ratio increases to 4.5 and 5 respectively as a result of Coulomb interactions attenuating the flux of protons. This figure shows that the relative importance of neutron- and proton-induced nuclear reactions varies throughout the atmosphere.

The production rate P of a cosmogenic nuclide j at atmospheric depth D is given by

$$P_j(D) = \sum_i N_i \sum_k \int_0^\infty \sigma_{ijk}(E_k) \cdot J_k(E_k, D) dE_k, \quad (10.3-1)$$

where the production is summed over interactions with all the i target nuclei (e.g., O, N, or Ar). This equation is derived from (B10.2.1-9) and shows that the production rate P is given by the product of the depth and latitude-dependent particle flux of the k th component of the nuclear cascade, J_k , the cross-section σ , and the density of the target atoms N_i . For N_i , the unit atoms per gram is used to

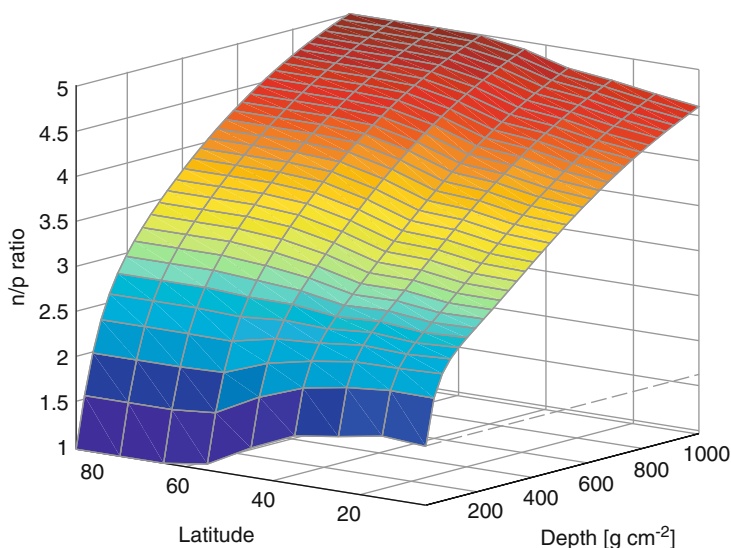


Fig. 10.3-2 Neutron/proton ratio for the integrated fluxes above 20 MeV as a function of atmospheric depth and geomagnetic latitude for the present magnetic field ($M = 1$), and solar modulation potential $\Phi = 550$ MeV

make it independent of the atmospheric depth. Integration over energy is necessary because both the particle fluxes and cross-sections are energy dependent. In practice, protons and neutrons are the only components of the nuclear cascade (Fig. 10.2.1-1) that make a significant contribution in the atmosphere. The difficulty in applying (10.3-1) is that the fluxes of the primary and secondary cosmic ray particles have to be known as a function of energy and atmospheric position, together with all the energy-dependent cross-sections that contribute to the production rate of the nuclide j .

As we have seen in Sect. 10.2.3, the particle fluxes can now be calculated using computer codes such as GEANT and FLUKA. Unfortunately, the situation in respect of the cross-sections is less favourable. The processes taking place during the interaction of a particle (n, p) with a nucleus are very complex and there are no nuclear models that provide accurate cross-sections over all energies of interest. Most models used for estimating the cross-sections are rough approximations with uncertainties of up to 100% (Rudstam 1966). The only reliable solution is, therefore, to determine cross-sections experimentally as described in the next section.

Once the neutron and proton fluxes and the cross-sections are known, the production rate is obtained by summing over all the targets that contribute to the production rate. The main and some rare target elements in the atmosphere and their concentration by weight are given in Table 10.1-1.

10.3.1 Early Production Models

As outlined above, the production rate of a cosmogenic nuclide is determined by the n and p fluxes and the corresponding cross-sections (see 10.3). Studies of the cosmogenic radionuclides were initiated in the middle of the twentieth century, and at that time, the particle spectra and their dependence on altitude, latitude, and solar and geomagnetic modulation were not known. Furthermore, cross-sections were poorly known. The pioneers in the field, therefore, chose a different approach (Lal and Peters 1967). They determined the rate of interaction between cosmic ray particles and nuclear matter using photographic emulsions exposed for several years at different altitudes and latitudes. By counting the number of interactions (stars) on a photo plate and multiplying it with the probability (yield function) that a certain cosmogenic nuclide would be produced, they obtained the production rates without a detailed knowledge of the physical processes involved. In the case of the short-lived cosmogenic radionuclides (e.g., ^7Be), the yield functions were determined experimentally. In the case of the long-lived nuclides (e.g., ^{10}Be), the total yield of the element in question was evaluated, and this yield was distributed among the different isotopes of the element using a semi-empirical equation (Lal and Peters 1967).

With time, several attempts were made to estimate the fluxes using mathematical solutions of the equations of transport (Blinov 1988). However, with a growing number of experimentally measured cross-sections and more powerful computers,

there was a shift to the use of the Monte Carlo technique, as described in Sect. 10.2.3. This approach has several advantages: (1) Particle spectra and cross-sections can be checked and improved independently from each other. (2) The production rate of other cosmogenic nuclides can be calculated easily provided the corresponding cross-sections are available. (3) Sensitivity studies can be carried out to estimate the effects of uncertainties in the particle fluxes and the cross-sections. There are now a number of comprehensive calculations based on these techniques (Masarik and Beer 1999, 2009; Webber and Higbie 2003; Usoskin and Kovaltsov 2008; Webber et al. 2007), and we present and discuss the results of one set of these calculations in the following sections.

10.3.1.1 BOX Nuclear Reactions

Nuclear reactions are collisions between particles or ions that lead to new particles or ions. If a high energy cosmic ray proton collides with an atom in the atmosphere, energy and momentum are transferred to the nucleus of the atom. A large part of the energy and momentum is carried away by protons, neutrons, and alpha particles expelled from the nucleus. In addition, new particles such as Kaons, Pions, and Muons are generated by the interaction (see Fig. 10.2.1-1). These secondary particles are responsible for the development of the hadronic cascade in the atmosphere which can be simulated by Monte Carlo techniques. If the energy of the primary cosmic ray particle is high enough (>1 GeV), its interaction with the target nucleus can be considered primarily as a nucleon–nucleon interaction of the incoming nucleon hitting a nucleon in the nucleus. The exact structure of the nucleus is only of minor importance. This direct nucleon–nucleon process is fast.

As a result of the inelastic interaction of the primary particle with the target nucleus, some energy and momentum reside in the nucleus after the fast direct reaction. The excited nucleus can break up into fragments, a process called spallation. This process is important for the formation of a large variety of cosmogenic radionuclides with masses smaller than that of the target nucleus. The following example describes the interaction of a proton with ^{14}N producing ^{10}Be : $^{14}\text{N}(p,1n4p)^{10}\text{Be}$. A high-energy proton hits a nitrogen atom and removes one neutron and 4 protons, including the primary one. The loss of these nucleons is only partly the result of the direct collision. In analogy to a water droplet which loses water molecules when heated, an excited nucleus can “evaporate” nucleons to get rid of its energy and momentum.

In reality, the nuclear processes are very complex and can only be described quantum mechanically. However, there are relatively simple models which reproduce the basic features qualitatively. Accurate quantitative values, e.g., for nuclear cross-sections have to be determined experimentally.

10.3.2 Production Cross-Sections

As discussed in Box 10.2.1.1, cross-sections give the probability that an interaction of a specific particle (e.g., a neutron) with a specific target (e.g., a nitrogen atom) will result in the production of a specific radionuclide (e.g., a ^{10}Be atom). As discussed in Sect. 10.3.1, most cross-sections are determined using experimental techniques in which a thin target of a specific element is irradiated with protons or neutrons. Then the cross-section can be calculated by measuring the number, N , of nuclides produced by a fluence F incident on a target of thickness $\times g\text{ cm}^{-2}$. F is the beam intensity integrated over the irradiation time monitored by directly measuring either the beam current or the yield of a standard nuclear reaction with a well-known cross-section. Then from (B10.2.1-9),

$$N = (L\sigma X/A)F \quad (10.3.2-1)$$

$$\sigma = NA/(LXF) \quad (10.3.2-2)$$

Proton cross-sections are comparatively easy to determine because accelerators provide high-intensity proton beams with well-defined energies. Since neutrons cannot be accelerated and deviated by electric and magnetic fields, neutron beams have to be generated using proton reactions such as $\text{Li}(p, n)$. As a consequence, intense monoenergetic neutron beams are rare and this leads to a general lack of neutron cross-sections. The main difference between proton and neutron cross-sections occurs at low energy because the positively charged proton is repelled by the protons in the target nucleus (Coulomb barrier), while the neutron can enter the nucleus at very low energies without any problems. A good example is the production of ^{14}C , which is almost exclusively generated in the atmosphere by thermal neutrons (having a typical energy of 0.025 eV corresponding to the energy of a molecule in air at room temperature) interacting with nitrogen. At very high energies (above 1 GeV), the charge effect vanishes, and proton and neutron cross-sections become identical.

A curve showing a cross-section as a function of energy is often called an excitation function. Typically, an excitation function starts at a threshold energy corresponding to the minimum energy required for this reaction to occur. After a relatively steep increase, a maximum is reached. A further increase in energy usually leads to a slight decrease due to the onset of additional competing reactions with higher thresholds. Finally, a steady state is reached and the cross-section becomes independent of energy.

To fix ideas, the cross-sections (i.e., the excitation functions) for the production of ^{10}Be and ^7Be are shown for the target elements N and O in Fig. 10.3.2-1, and for the production of ^{36}Cl from Ar in Fig. 10.3.2-2.

The following nuclear reactions are involved in the production of ^{10}Be in the atmosphere:

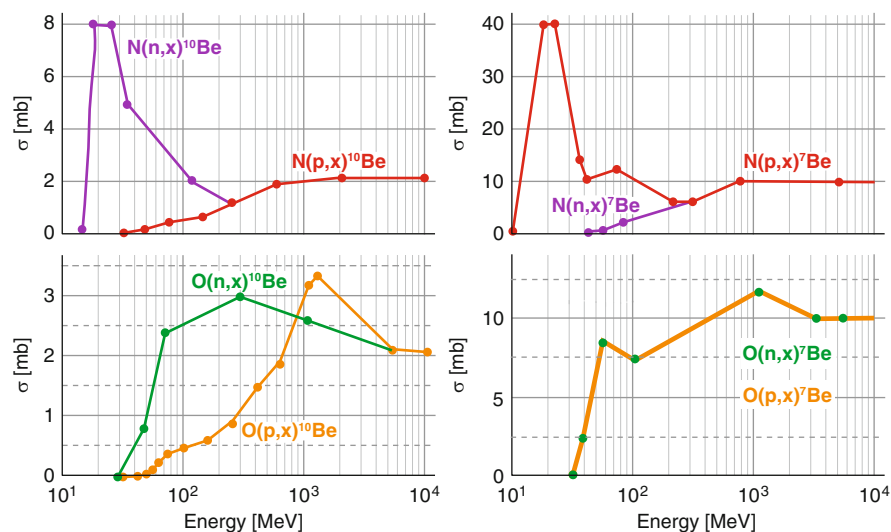


Fig. 10.3.2-1 Cross-sections for ^{10}Be and ^7Be from nitrogen (*upper panel*) and oxygen (*lower panel*). The thresholds for the reactions are between 20 and 30 MeV for ^{10}Be and 10 MeV for ^7Be . The largest values are 8 mb (1 mb = 10^{-3} b = 10^{-27} cm 2) in the case of ^{10}Be and 40 mb in the case of ^7Be . While all cross-sections for ^{10}Be approach 2 mb at high energies, they approximate 10 mb for ^7Be . These differences clearly indicate that the production rate of ^7Be is larger than that of ^{10}Be (see Sect. 10.3.3)

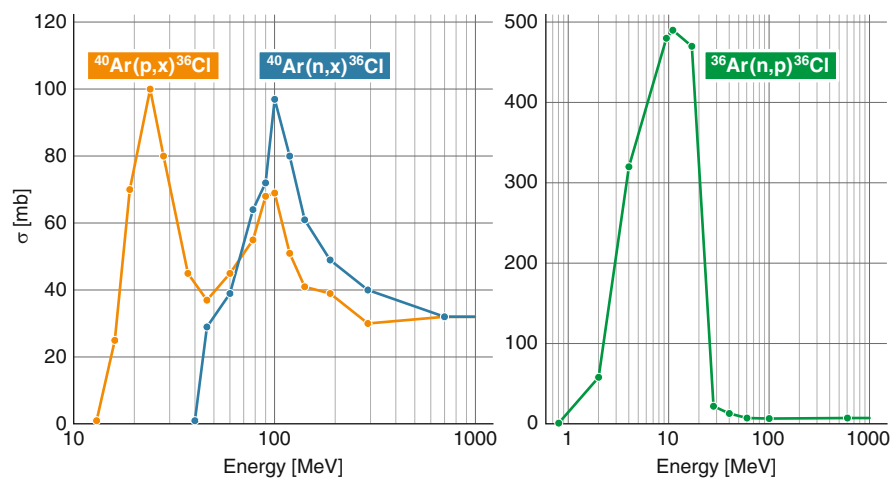
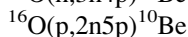
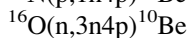
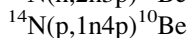
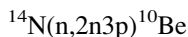


Fig. 10.3.2-2 Production cross-sections for ^{36}Cl from ^{40}Ar and ^{36}Ar . Although the abundance of ^{36}Ar is only 0.34%, the cross-section at 10 MeV is very large and, therefore, cannot be neglected



This nomenclature has the following meaning. For the first reaction, one neutron entering a ^{14}N nucleus must expel two neutrons and three protons to turn it into a ^{10}Be nucleus. This can be done in different ways, e.g., by emitting an alpha particle plus a proton.

Figure 10.3.2-1 shows that the thresholds for the ^{10}Be production from nitrogen are about 15 MeV for neutrons and 33 MeV for protons. In the case of the ^{10}Be production from oxygen, the threshold for both nucleons is about 30 MeV. As we will discuss later, Fig. 10.3.2-1 indicates that the neutron reaction with nitrogen is the most important for the production of ^{10}Be . At energies above a few hundred MeV, all the cross-sections approach a constant value.

As summarized by (10.3-1), the production rate for a given radionuclide is obtained by summing over all the targets that contribute to the production rate. The main and some rare target elements in the atmosphere and their concentration by weight are given in Table 10.1-1 (atmospheric composition). As already mentioned earlier, the main atmospheric target elements for the production of ^7Be , ^{10}Be , and ^{14}C are N and O, and Ar for ^{36}Cl and ^{26}Al .

Figure 10.3.2-2 displays the excitation functions for the production of ^{36}Cl through the three reactions $^{40}\text{Ar}(p,2p3n)^{36}\text{Cl}$, $^{40}\text{Ar}(n,1p4n)^{36}\text{Cl}$, and $^{36}\text{Ar}(n,p)^{36}\text{Cl}$. Note the high values of the cross-sections (120–400 mb) compared to those that generate ^7Be and ^{10}Be . The low abundance of Ar (~1%) in the atmosphere is partly compensated by the comparatively large cross-sections and low thresholds.

As a result of the different thresholds and shapes of the cross-sections, the production rates depend on the “hardness” of the n and p differential energy spectra. This leads to somewhat different production rates as a function of atmospheric depth and geomagnetic latitude (see Sect. 10.3.3).

To this point, we have discussed the production of ^7Be , ^{10}Be , and ^{36}Cl by neutrons and protons with energies that are greater than the nuclear binding energy (>8 MeV). By way of contrast, the other very important cosmogenic nuclide, ^{14}C , is primarily (99%) produced by thermal neutrons through the $^{14}\text{N}(n,p)^{14}\text{C}$ process. In this case, the production has been simulated using the MCNP code which used the simulated neutron spectra of GEANT as input and continued the simulation of the cascade until thermal energies were reached. Based on these thermal neutron fluxes, the ^{14}C production rates were computed as a function of atmospheric depth, latitude, modulation function (Φ), and magnetic dipole strength (M) in a manner similar to that outlined above (Masarik and Beer 1999).

10.3.3 Production Rates and Inventories

In this section, we calculate the production rate for ^{10}Be and ^7Be as a function of atmospheric depth, geomagnetic latitude, solar modulation, and geomagnetic

modulation. Based on these production rates, the global inventories (the total amount of these radionuclides under steady state conditions) are calculated. We have chosen ^{10}Be and ^7Be to discuss the characteristic properties that apply, to a greater or lesser extent, to all the cosmogenic nuclides produced in the atmosphere. All the other cosmogenic radionuclides show a qualitatively similar behaviour but differ considerably as far as their absolute production rates are concerned. To demonstrate the relatively small relative differences that do occur, we will discuss the manner in which $^{10}\text{Be}/^7\text{Be}$ and the $^{36}\text{Cl}/^{10}\text{Be}$ ratios depend upon latitude and atmospheric depth.

All the following calculations use (10.3-1) and are based on the proton and neutron flux simulation carried out by Masarik and Beer (1999, 2009). The general features of those fluxes have been discussed in Sect. 10.2.3. As mentioned previously, the largest uncertainty in these calculations is due to a lack of well-constrained cross-sections, particularly for the neutron reactions. However, whenever better cross-sections become available, the calculations can be repeated, leading to improved estimates of the production rates.

To this point, all of our discussions have been for cosmic ray protons alone. We now consider the contributions made by the alpha and heavier components of the cosmic radiation (Fig. 5.4-4). Section 5.8 has discussed the fact that for a given energy per nucleon, the alphas and heavies are able to penetrate the geomagnetic field more easily than does a cosmic ray proton. For example, the discussion associated with (5.8.2-1) shows that a proton must have $E > 14$ GeV to reach the equatorial regions, while alphas and heavies can do so if their energy per nucleon is >7 GeV. The solar modulation increases rapidly towards lower energies, and the percentage variation in the cosmic ray flux at a cut-off energy of 7 GeV is approximately twice that at 14 GeV. About 40–49% of the production of the cosmogenic nuclides is due to the alphas and heavies, and it is therefore clear that we will underestimate the solar (and magnetic) modulation effects unless we take this difference into account (McCracken 2004).

The manner in which this allowance is made is summarized in Box 10.3.3.1. This procedure has been applied in several different ways in the literature after 2004 (Masarik and Beer 2009; McCracken et al. 2004; Webber et al. 2007) and has been used in the derivation of all the figures and tables in this section. That is, from this point onwards, all the results include the effects of the full cosmic ray spectrum, including protons, alphas, and heavier nuclei.

10.3.3.1 BOX Treating Non-proton Particles

To date, inclusion of the contributions made by the alpha and heavy components of the cosmic radiation has been made after computation of the the production rate for protons alone. This was necessary to eliminate the additional computer time needed if the same particle by particle approach were to be used to compute the proton and neutron fluxes as outlined

in Sect. 10.2.3. The following outlines the way in which the proton results are used to compute the production for all components of the cosmic radiation.

For proton cosmic rays, let $P_j^p(E_p, \Phi, M)$ be the production rate of the j th radionuclide at a proton cut-off energy of E_p , modulation function Φ , and dipole moment M , for a proton flux J . Let the fluxes of the proton, alpha, and heavy components bear the ratios r_p , r_a , and r_m to the total cosmic ray flux J (these ratios are typically 0.87, 0.12, and 0.01). Assuming a mean value for the atomic mass numbers of the heavy component to be 12, the number of nucleons associated with the alphas and heavies is $(4 r_a + 12 r_m) J$. In a first approximation at high energies, each of these nucleons generates a nuclear cascade similar to that produced by a cosmic ray proton. The alpha and heavy cosmic rays have a cut-off energy which is approximately half of that of the protons [see Eq. (5.3-7b) and Fig. 5.3-1]. The production rate for the alpha and heavy components is, therefore,

$J \cdot [4 r_a + 12 r_m] \cdot P_j^p(E_p/2, \Phi, M)$, while that due to the cosmic ray protons is $J \cdot r_p \cdot P_j^p(E_p, \Phi, M)$. Knowing $P_j^p(E_p, \Phi, M)$ as a function of E_p , the production of the j th cosmogenic radionuclide at proton cut-off energy E_p is

$$P_{\text{total}}(E_p, \Phi, M) = r_p \cdot P_j^p(E_p, \Phi, M) + [4r_a + 12r_m] \cdot P_j^p(E_p/2, \Phi, M) \quad (\text{B10.3.1.1-1})$$

The production rate of the j th radionuclide due to cosmic ray protons was originally evaluated versus geomagnetic latitude using the procedures given in Sects. 10.3.1 and 10.3.2. Using these curves, and the above equation B10.3.1.1-1 for $P_{\text{total}}(R_p, \Phi, M)$, the total production rate for protons, alphas, and heavies can be calculated. McCracken (2004) used this procedure for all latitudes, while Masarik and Beer (2009) evaluated the multiplicative factor to be applied to the proton production at a number of discrete rigidities, and then used interpolated values to estimate the multiplicative factors at all latitudes. The two approaches are equivalent.

Figure 10.3.3-1 shows the depth dependence of the ^{10}Be production rate in atoms per gram of air and per second for the present geomagnetic field ($M = 1$) and the solar modulation function $\Phi = 550$ MeV. The production rate is highest at high geomagnetic latitudes and decreases monotonically with increasing atmospheric depth by more than a factor of 1,000. As a result of the increasing geomagnetic cut-off rigidity (i.e., the shielding of the primary cosmic rays by the geomagnetic field), the production rate at the top of the atmosphere is reduced by a factor of almost 10 between the poles and the equator. This latitude effect is reduced deeper in the atmosphere because only the secondaries induced by very high-energy primaries can penetrate deep into the atmosphere, and their flux is much less influenced by the geomagnetic rigidity. Another effect of the high cut-off energy at low latitudes is

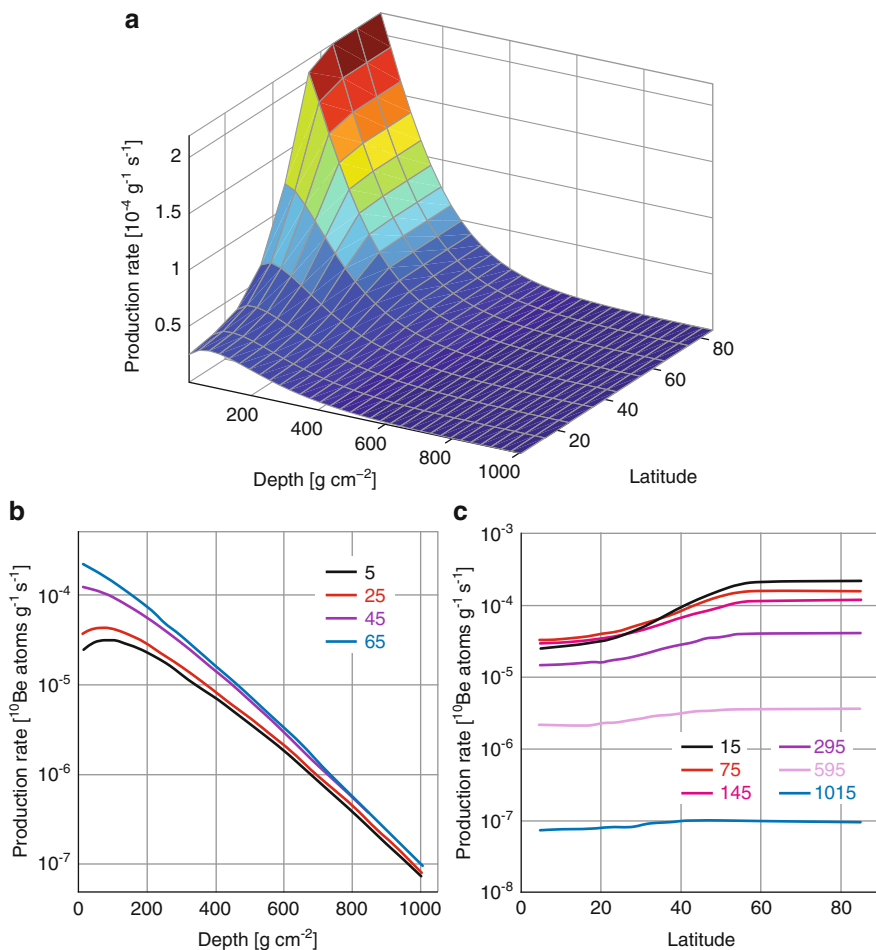


Fig. 10.3.3-1 Dependence of the ^{10}Be production rate in atoms per gram of air and per second on atmospheric depth and geomagnetic latitude. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations. The figure consists of a 3-dim plot (a) with linear scales and the two corresponding projections (b) and (c) with semi-logarithmic scales. The geomagnetic field intensity M is 1 corresponding to the present field and the solar modulation function Φ is set to 550 MeV

that the flux of low-energy secondary protons increases from the top of the atmosphere down to $\sim 100 \text{ g cm}^{-2}$ as the cascade develops. As a consequence, the ^{10}Be production rate exhibits a peak at $\sim 100 \text{ g cm}^{-2}$.

Figure 10.3.3-2 examines the effect of changes in the solar modulation Φ upon the ^{10}Be production rate. The magnetic field is set to the present value ($M = 1$) and the production rate is integrated over the full atmospheric depth. Panel b shows that the modulation effect is small for latitudes $< 30^\circ$ and constant above 60° . In the case of no solar activity ($\Phi = 0 \text{ MeV}$), the latitudinal effect is a factor of 7, while

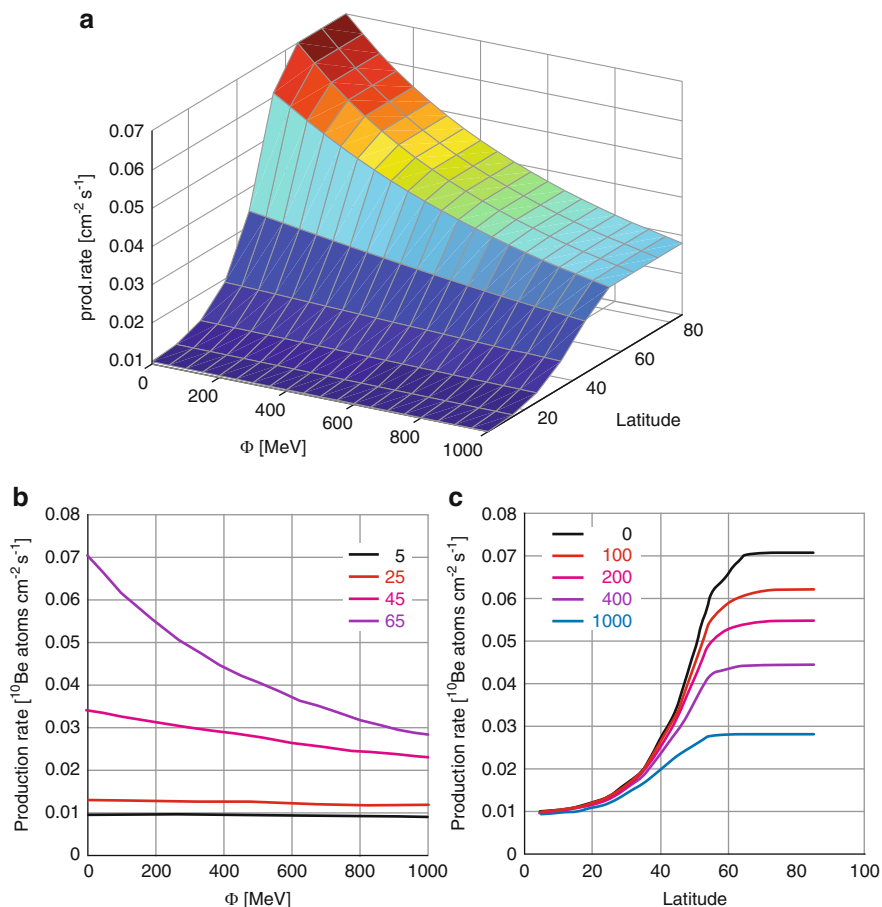


Fig. 10.3.3-2 Dependence of the ^{10}Be production rate upon geomagnetic latitude and the solar modulation function Φ for the present-day geomagnetic field ($M = 1$). The production rate is integrated over the full atmospheric depth. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations. The projections of the 3-dimensional plot (a) on the Φ and latitude axis are shown in the plots (b) and (c) respectively

for strong solar activity ($\Phi = 1,000$ MeV), the latitudinal effect is reduced to a factor of about 3.

Figure 10.3.3-3 shows how the ^{10}Be production rate depends on the geomagnetic dipole moment (relative to the present) and the geomagnetic latitude. The latitude dependence upon dipole moment is strongly nonlinear. In Fig. 10.3.3-3b note that the “magnetic modulation” is strongest in equatorial regions, while it is non-existent at geomagnetic latitudes $>70^\circ$. This leads to an overlap in Fig. 10.3.3-3c of all curves at magnetic latitudes $>65^\circ$. If the geomagnetic field is switched off ($M = 0$), as may occur during a geomagnetic reversal, the production rate is constant over the globe. Note however that the persistence of non-dipole terms in the geomagnetic field at such times will result in some regional magnetic modulation.

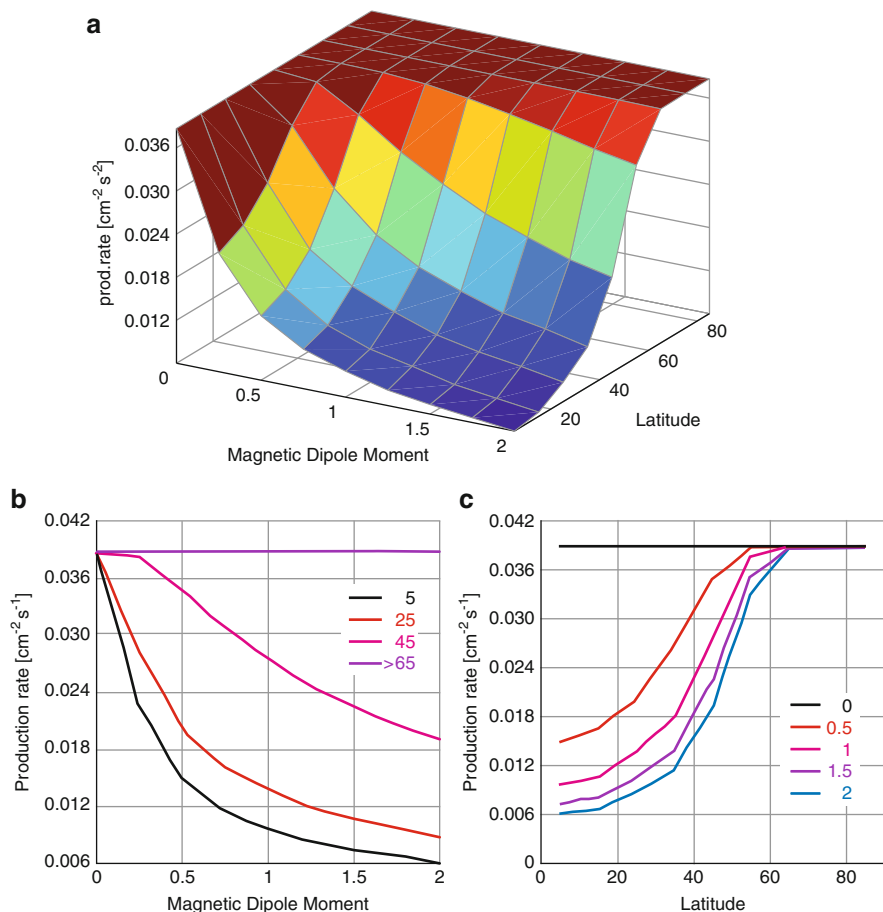


Fig. 10.3.3-3 Dependence of the ^{10}Be production rate on the relative geomagnetic dipole moment and the geomagnetic latitude for $\Phi = 550$ MeV. The production rate is integrated over all atmospheric depths. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations

Finally, Figure 10.3.3-4 shows how the mean global ^{10}Be production rate depends upon solar modulation and the strength of the geomagnetic dipole. First of all, it shows that the mean global production rate has a dynamic range of almost an order of magnitude from a maximum when the geomagnetic field is zero and the Sun quiet ($\Phi = 0$ MeV) to a minimum when the magnetic field is doubled and the Sun very active. As M increases, so the geomagnetic cut-off energy increases, removing the lower energy cosmic rays that suffer the strongest solar modulation. Thus, the amplitude of the solar modulation is a relatively strong function of strength of the geomagnetic dipole.

So far we have discussed the dependence of the ^{10}Be production rate upon atmospheric depth, geomagnetic latitude, solar activity, and geomagnetic field

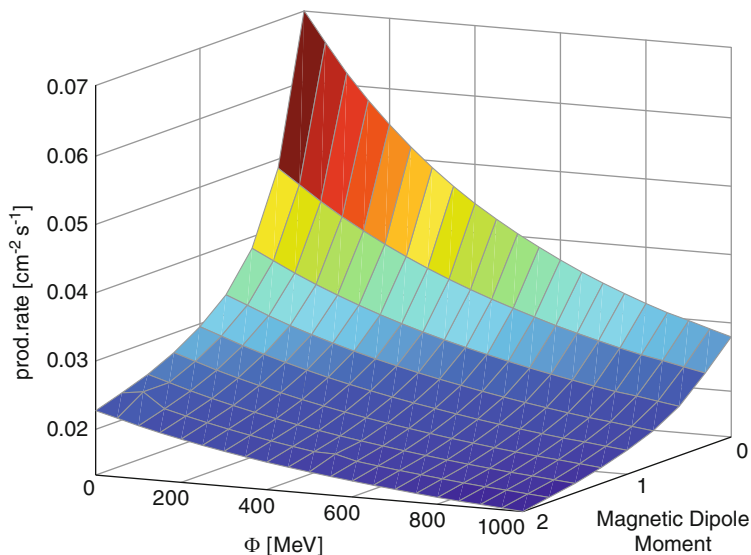


Fig. 10.3.3-4 Dependence of the mean global production rate of ^{10}Be upon the strength of the geomagnetic dipole, and solar modulation. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations

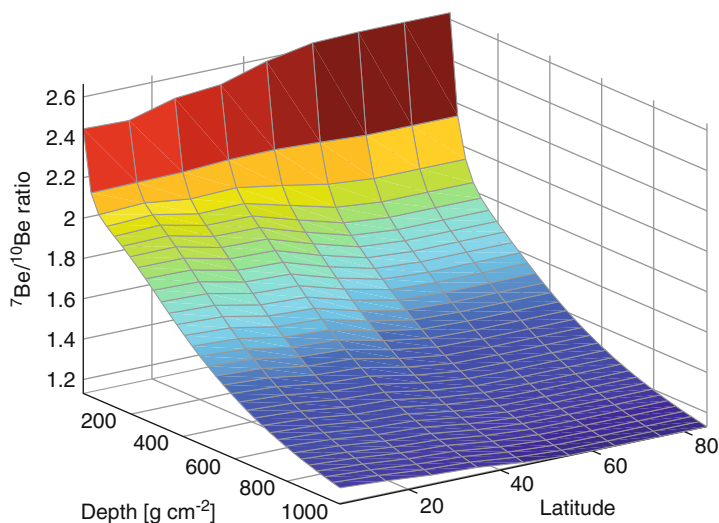


Fig. 10.3.3-5 $^7\text{Be}/^{10}\text{Be}$ production ratio for the present geomagnetic field and a solar activity of $\Phi = 550$ MeV. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations. As result of the different cross-sections and the different depth dependence of proton and neutron fluxes, the ratio decreases by about a factor of 2 from the top of the atmosphere down to sea level

intensity. Somewhat similar dependencies are observed for all the cosmogenic radionuclides. However, close inspection shows that there are some significant differences in the shapes of the specific production plots as a result of the different cross-sections, and the different depth dependencies of the proton and neutron fluxes. To demonstrate these differences in greater detail, we now examine the behaviour of several isotopic ratios. In the following, we investigate the $^7\text{Be}/^{10}\text{Be}$ and the $^{10}\text{Be}/^{36}\text{Cl}$ ratios which illustrate some key points, and which also play an important role in studies of atmospheric transport and in dating very old ice respectively (see Chaps. 13 and 23) (see Fig. 10.3.3-5).

The production rate of ^7Be is dominated below 500 MeV by proton-induced spallation of nitrogen (see Fig. 10.3.2-1), while in the case of ^{10}Be , the neutron-induced reactions are more important. In fact, for the present geomagnetic field and a solar activity of $\Phi = 550$ MeV, 76% of ^7Be and only 13% of ^{10}Be are produced by protons. Figure 10.3-2 shows that the proton flux decreases more rapidly than the neutron flux with increasing atmospheric depth, particularly so in the topmost 300 g cm $^{-2}$. As a consequence, there is a rapid decrease in the $^7\text{Be}/^{10}\text{Be}$ production ratio at the top of the atmosphere, followed by a steadier decline at lower altitudes. Overall, the ratio changes by a factor of about 2 between the top of the atmosphere and sea level. For high geomagnetic latitudes, the effect is slightly larger than that for low latitudes. The mean global production ratio is 1.54.

Figure 10.3.3-6 displays the case of the $^{10}\text{Be}/^{36}\text{Cl}$ production ratio for the present geomagnetic field and a solar activity of $\Phi = 550$ MeV. ^{36}Cl has a large proton

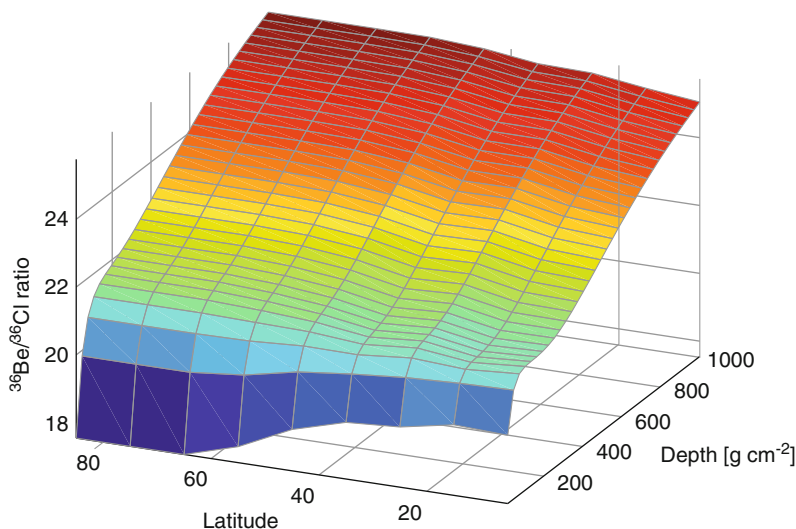


Fig. 10.3.3-6 $^{10}\text{Be}/^{36}\text{Cl}$ production ratio for the present geomagnetic field and a solar activity of $\Phi = 550$ MeV. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations. As result of the different cross-sections and the different depth dependence of proton and neutron fluxes, the ratio increases by about 40% from the top of the atmosphere down to sea level

cross-section for the reaction $^{40}\text{Ar}(2p3n)^{36}\text{Cl}$ (see Fig. 10.3.2-2), and this is responsible for the observed increase in the production ratio with depth. However, the protons contribute only about 32% to the total production rate. The change in the ratio is about 20% at low and 30% at high magnetic latitude. The mean global production ratio is ~ 20 .

Throughout the previous sections, we have repeatedly pointed out that the cosmogenic radionuclides are mainly produced by the neutrons generated in the nucleonic cascade initiated by a primary cosmic ray. Figure 10.3.3-7 illustrates

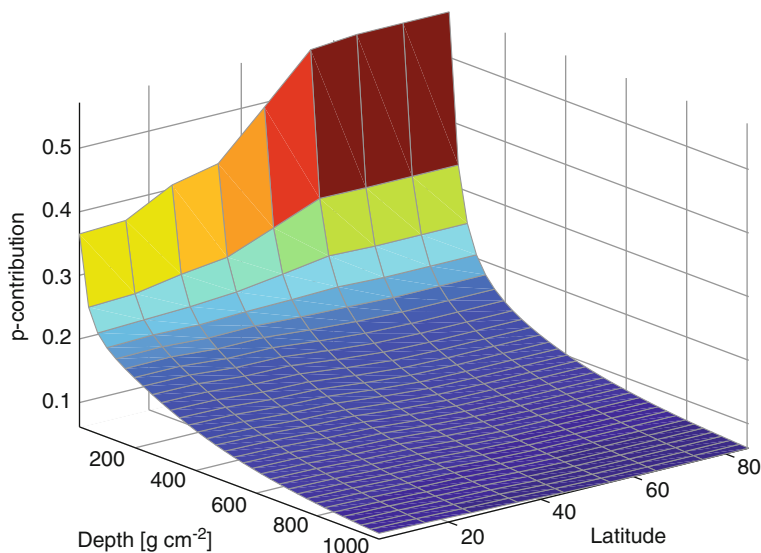


Fig. 10.3.3-7 Relative contribution of protons to the production of ^{10}Be as a function of atmospheric depth and geomagnetic latitude for the present geomagnetic field intensity and a solar activity of $\Phi = 550$ MeV

Table 10.3.3-1 Global inventory and activity of some cosmogenic radionuclides for steady state conditions. The protons, alpha particles, and heavy components of the cosmic ray flux are all included in these calculations. The mean global production rate is calculated for the present geomagnetic field intensity and a solar activity of $\Phi = 550$ MeV. Note that globally only about 16 g of ^7Be exists, while the global inventory of the much longer lived ^{10}Be amounts to 120 tons

Nuclide	Half-life (y)	Production rate (atoms $\text{cm}^{-2} \times \text{s}^{-1}$)	Global activity (10^{15} Bq)	Global inventory (g)
^3H	12.34	0.22/0.32		
^7Be	0.146	0.040	200	16
^{10}Be	1.387×10^6	0.021	110	120×10^6
^{14}C	5,730	2.02	10,000	62×10^6
^{26}Al	7.2×10^5	4.7×10^{-5}	0.24	0.34×10^6
^{36}Cl	3.01×10^5	1.12×10^{-4}	5.0	4.1×10^6
^{129}I	15.7×10^6	4.2×10^{-6}	0.00168	2.5×10^5

this fact using as an example the ^{10}Be production for the present geomagnetic field intensity and solar modulation function $\Phi = 550$ MeV.

While at the top of the atmosphere the protons contribute on average about 50%, their contributions drops to less than 10% at sea level. The latitude dependence is comparatively weak, with a larger p contribution at the poles at high altitude and a smaller one at low altitudes.

We finish this section about the production of cosmogenic radionuclides by estimating the global inventories and the total global activity of some of the commonly used nuclides. For this purpose, the assumption is made that steady state conditions are reached, meaning that production and decay are in equilibrium for the present geomagnetic field intensity and the solar modulation $\Phi = 550$ MeV. Table 10.3.3-1 shows that the inventory depends strongly on the half-life and ranges from 16 g (^7Be) to 120 tons (^{10}Be).

10.4 Production Results and Analytical Tools

The foregoing sections have examined the basic physics of the production of all the cosmogenic nuclides, and used ^{10}Be to illustrate the several dependencies that are of practical importance. We now provide a series of production graphs, followed by a brief discussion of analytical tools derived from these and other data (Figs. 10.4-1–10.4-3).

Section 6.5 introduced the concepts of the cosmic ray “specific yield” and “response” functions. These are used in the quantitative comparison of the data obtained with different cosmic ray detectors such as neutron monitors, and ionization chambers. Examples of several specific yield functions are given in Fig. 6.5-1. Using them, the relative sensitivity of several cosmic ray detectors to solar modulation is given in Fig. 6.6-1.

Clearly, the cosmogenic nuclides can be regarded as yet another type of cosmic ray detector, with the advantage that a data record exists that stretches many millennia into the past. For this reason, we now briefly review the specific yield and response functions that have been computed using Monte Carlo techniques at this time.

Figure 10.4-4 presents an early estimate of the response function of ^{10}Be . Examination of (6.5-2) shows that the response function of a cosmic ray detector can be obtained by differentiation of the latitude dependence of its output. In an analogous manner, Fig. 10.4-4 was obtained by differentiation of the results of the GEANT calculations that provided a latitude curve of the production of ^{10}Be similar to those in Figure 10.4-1 (McCracken 2004).

Note in particular that ^{10}Be has a peak energy sensitivity well below that of a neutron monitor. Thus, the neutron monitor response peaks at ~ 4.5 GeV, while for $\Phi = 600$ MeV (present day), the ^{10}Be peaks at 1.8 GeV. Equation (6.5-2) shows that the response function depends directly upon the nature of the cosmic ray spectrum, and that consequently, the peak energy of response depends on the

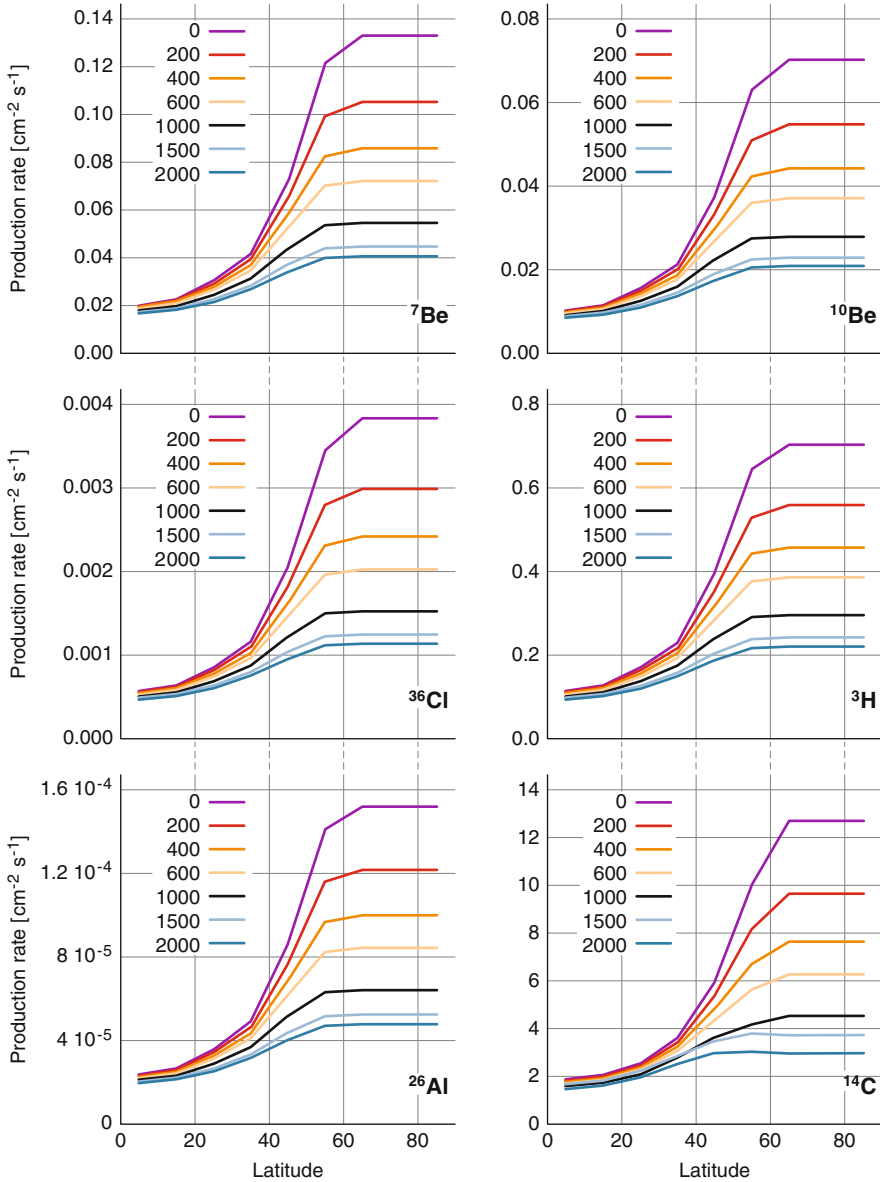


Fig. 10.4-1 The latitude dependence of the depth-integrated production of ${}^3\text{H}$, ${}^7\text{Be}$, ${}^{10}\text{Be}$, ${}^{14}\text{C}$, ${}^{26}\text{Al}$, and ${}^{36}\text{Cl}$ for different solar modulation functions Φ and the present magnetic field ($M = 1$). The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations

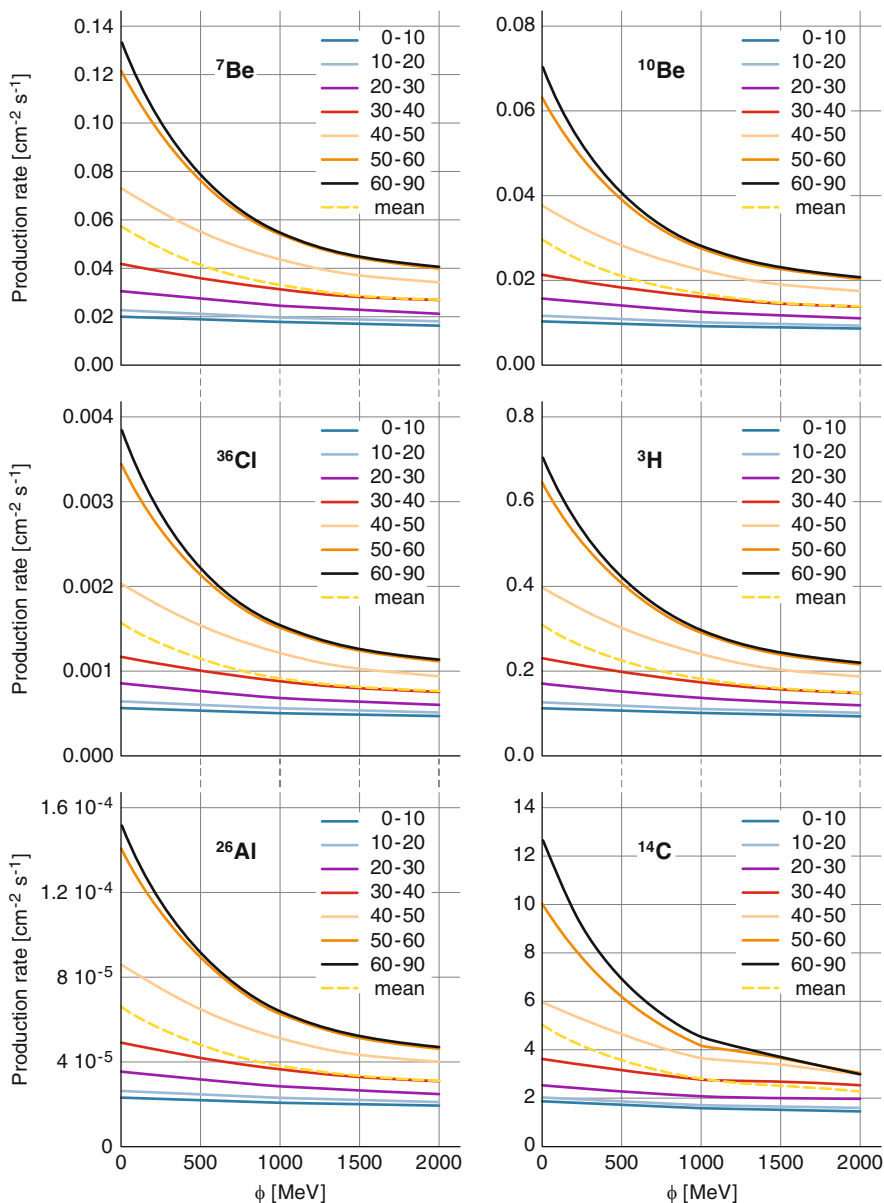


Fig. 10.4-2 The dependence of the depth-integrated production of ^3H , ^7Be , ^{10}Be , ^{14}C , ^{26}Al , and ^{36}Cl upon the modulation function Φ , for nine different bands of geomagnetic latitude, for the present-day magnetic dipole strength. The dotted yellow line depicts the global average production. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations

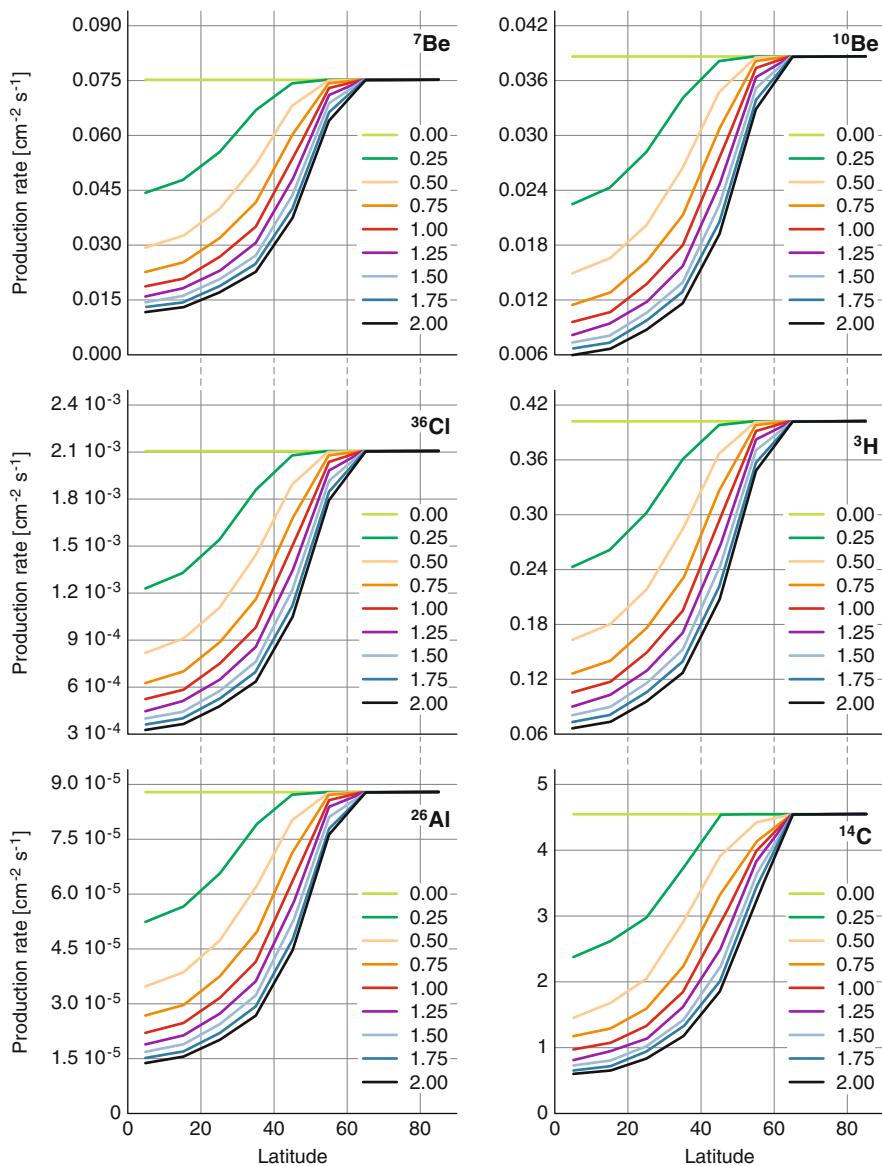
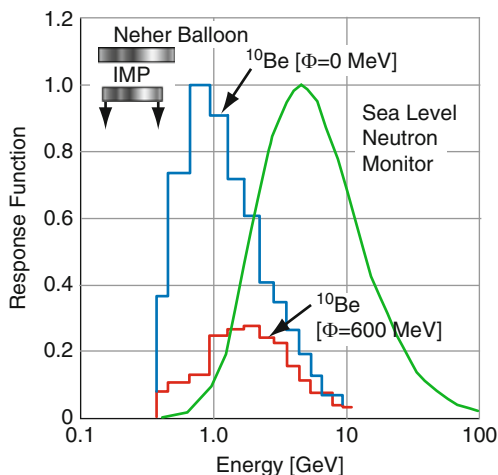


Fig. 10.4-3 The dependence of the depth-integrated production of ${}^3\text{H}$, ${}^7\text{Be}$, ${}^{10}\text{Be}$, ${}^{14}\text{C}$, ${}^{26}\text{Al}$, and ${}^{36}\text{Cl}$ upon latitude, and the strength of the geomagnetic dipole relative to the present-day value, for $\Phi = 550$ MeV. The proton, alpha particle, and heavy components of cosmic ray flux are all included in these calculations

Fig. 10.4-4 Comparison of the response functions of a sea-level neutron monitor, and ^{10}Be for two different levels of solar modulation, $\Phi = 0$ and 600 MV (McCracken 2004)



level of solar modulation. As an example, Figure 10.4-4 shows that the peak energy of the ^{10}Be response shifts down to 0.8 GeV for $\Phi = 0$ MeV. The amplitude of the solar modulation varies roughly as the reciprocal of cosmic ray rigidity, so converting the peak energies of 1.8 and 4.5 GeV to rigidity indicates that the solar modulation will be $\sim(5.4/2.6) = 2.1$ greater in the ^{10}Be than in the neutron monitor data over the past 50 years. This is in good agreement with observation.

The nucleonic cascade models such as those used in this chapter are now used to compute the specific yield functions of the cosmogenic nuclides directly. The procedure is very similar to that used in Sect. 10.2. Some millions of monoenergetic cosmic rays are “fired” at the model atmosphere, yielding a modelled nucleonic cascade, and the cosmogenic production corresponding to the incident cosmic ray energy.

Figure 8.2.3-4 presents the results of such a calculation (Webber et al. 2007). Because the cosmogenic yields are very low at $E = 20$ MeV (the lowest energy used), 10^7 protons were used at each energy. Note in particular the “bumps” in the ^7Be and ^{36}Cl response curves in the range 20–60 MeV. These are the consequence of the high values of the cross-sections for these reactions given in Figs. 10.3.2-1 and 10.3.2-2 and discussed in connection with Figs. 10.3.3-5 and 10.3.3-6.

References

- Blinov A (1988) The dependence of cosmogenic isotope production rate on solar activity and geomagnetic field variations. In: Stephenson FR, Wolfendale AW (eds) *Secular solar and geomagnetic variations in the last 10,000 Years*. Kluwer, Durham, pp 329–340
- Brun B, et al. (1987) *GEANT3 User's guide Rep.*, 584 pp, Eur Org for Nucl Res, Geneva
- Castagnoli G, Lal D (1980) Solar modulation effects in terrestrial production of C-14. *Radiocarbon* 22(2):133–158

- Lal D, Peters B (1967) Cosmic ray produced radioactivity on the Earth. In: Flüge S (ed) *Handbuch für Physik*. Springer, Berlin, pp 551–612
- Masarik J, Beer J (1999) Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res* 104(10):12099–12111
- Masarik J, Beer J (2009) An updated simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res-Atmos* 114:D11103
- McCracken KG (2004) Geomagnetic and atmospheric effects upon the cosmogenic Be-10 observed in polar ice. *J Geophys Res-Space Phys* 109(A4):4101
- McCracken KG, McDonald FB, Beer J, Raisbeck G, Yiou F (2004) A phenomenological study of the long-term cosmic ray modulation, 850–1958 AD. *J Geophys Res-Space Phys* 109:A12103
- Rudstam G (1966) Systematics of spallation yields. *Z Naturforsch* 21a:1027–1041
- Simpson JA (2000) The cosmic ray nucleonic component: the invention and scientific uses of the neutron monitor (Keynote lecture). *Space Sci Rev* 93(1–2):11–32
- Steinhilber F, Abreu JA, Beer J (2008) Solar modulation during the Holocene. *Astrophys Space Sci Trans* 4:1–6
- Usoskin IG, Kovaltsov GA (2008) Production of cosmogenic Be-7 isotope in the atmosphere: full 3-D modeling. *J Geophys Res-Atmos* 113:D12107
- Webber WR, Higbie PR (2003) Production of cosmogenic Be nuclei in the Earth's atmosphere by cosmic rays: Its dependence on solar modulation and the interstellar cosmic ray spectrum. *J Geophys Res* 108(9):1355–1365
- Webber WR, Higbie PR, McCracken KG (2007) Production of the cosmogenic isotopes ^3H , ^7Be , ^{10}Be , and ^{36}Cl in the earth's atmosphere by solar and galactic cosmic rays. *J Geophys Res - Part A - Space Phys* A10106:10107

Chapter 11

Production of Cosmogenic Radionuclides in Other Environmental Systems

11.1 Introduction

The atmosphere is the main source of the cosmogenic radionuclides simply because that is where the cosmic radiation first encounters terrestrial matter, and also because it is thick enough that few of the secondary particles in the nucleonic cascade (Sect. 10.1) reach the surface of Earth (e.g. Fig. 10.2.3-3). Accordingly, the production in the atmosphere has been discussed in detail. However, this does not mean that production does not take place in other environmental systems, or that this production is not important. On the contrary, in situ production, as it is often called, plays an increasingly important role in geology and geomorphology and the number of applications is steadily increasing (see Chap. 21). Production in meteorites and lunar material also provides important information regarding the Sun and the structure of our galaxy.

There are three main differences between in situ cosmogenic production compared to that in the atmosphere. (1) The incident cosmic ray spectrum is different. (2) There is a richer mix of elements in terrestrial and extraterrestrial matter, and some important target nuclei are much more abundant than in the atmosphere. The production of ^{36}Cl is a good example. While it is produced in the atmosphere from Ar which is relatively rare, in the lithosphere it is produced from the relatively common nuclide ^{40}Ca . (3) Generally, the cosmogenic radionuclides remain fixed at the point where they are produced. This is a crucial difference, without which some important applications would not exist. The main exceptions are the cosmogenic noble gases which may diffuse over long periods of time depending on the crystal structure of the solid body. Of these, ^3He and ^4He are the most mobile stable isotopes.

The principles of the production of cosmogenic radionuclides in solid matter are discussed in Box 11.1.1. Note, in particular, that we must allow for the fact that the surface of the solid body is eroded with time, as in the case of rocks exposed to weathering, or that additional absorbing material may be added to the surface over time, as in the case of an ice sheet or a sediment in a shallow lake.

In Sect. 11.2 we then discuss in situ production underground, including some additional production processes that are important for specific applications such as ground water dating. Finally we address in situ production in extraterrestrial matter in Sect. 11.3. Extraterrestrial matter is exposed to the unshielded cosmic ray intensity and therefore the production rates are large and the detection of cosmogenic radionuclides is comparatively easy. This is particularly so for the noble gases and as a consequence there is a long history of measuring them to determine the exposure ages of meteorites (see Sect. 23.2.2).

11.1.1 BOX Production of a Cosmogenic Radionuclide in Solid Matter

Box 10.2.1.1 of Chap. 10 derived the production rate of a cosmogenic nuclide in terms of the incident nucleonic flux, and the cross-section of the reaction in question. We now apply those concepts for the determination of the concentration of a cosmogenic radionuclide in terrestrial or meteoric matter, where the surface is being eroded (or augmented) as a function of time.

From (B10.2.1-9) we can show that the instantaneous production rate at depth x (cm) depends on the production rate at the surface (written P_0) and decreases exponentially with $\rho x/\Lambda$ (see Fig. 11.1-1). Thus writing $\rho x = X$, the depth in g cm^{-2} as discussed in Box 10.2.1.1, the production rate is given by

$$P(X) = P_0 e^{-x\rho/\Lambda} = P_0 e^{-X/\Lambda} \quad (11.1-1)$$

Clearly, this results in a gradient in the concentration of the nuclide, dN/dX , the concentration decreasing with increasing depth in g cm^{-2} .

Consider now the case where there is steady erosion of the surface at a rate of q ($\text{g cm}^{-2} \text{ s}^{-1}$). This means that the infinitesimal thickness dX is moving downwards at rate q in order that it should remain a distance x below the

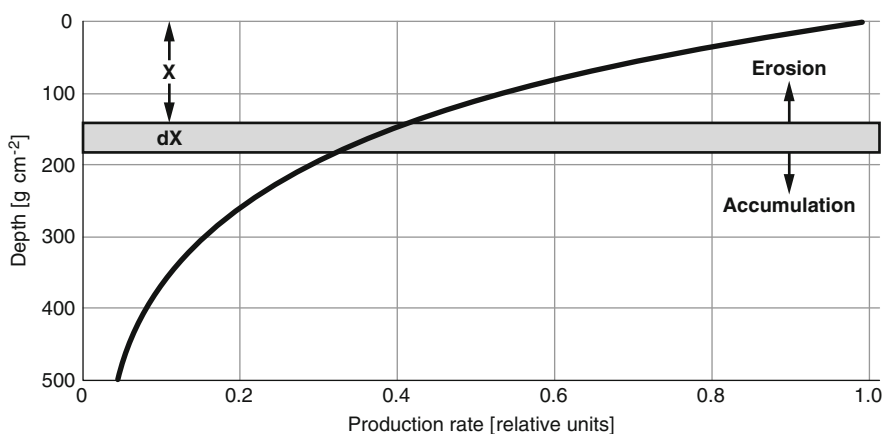


Fig. 11.1-1 Dependence of the production rate in solid matter on the depth assuming erosion or accumulation at the surface. Note that the depth is given in g cm^{-2}

surface. In other words, the concentration in dX is changing for three reasons: (1) the “new” production $P_0 e^{X/\Lambda}$ results in an increase; (2) dX has moved deeper, and now contains material that has experienced less production in the past, which results in an apparent decrease; and (3) the radionuclide is decaying with the decay constant λ , also resulting in a decrease in concentration. These three contributions are written as the differential equation

$$\frac{dN(X, t)}{dt} = P_0 e^{-X/\Lambda} + q \frac{dN(X, t)}{dX} - \lambda N(X, t) \quad (11.1-2)$$

P_0 : production rate at the surface in atoms $\text{g}^{-1} \text{s}^{-1}$

X : depth in g cm^{-2}

Λ : attenuation mean free path in g cm^{-2}

q : change in depth of layer dX in $\text{g cm}^{-2} \text{s}^{-1}$

λ : radioactive decay constant in s^{-1}

As defined above, $q > 0$ for erosion at the surface, while $dN(X, t)/dX$ is negative.

Exactly the same argument applies in the case where extra material is added to the surface over time. Common examples are where the surface is being buried by new layers of sediment, or by snow in the case of an ice sheet. Consideration will show that the same differential equation applies, except now q is negative (accumulation corresponds to negative erosion). Note that in the literature, $N(X, t)$ is often written as $C(X, t)$.

If P_0 and q are constant the solution of (11.1-2) is given by:

$$N(X, t) = \frac{P_0}{\lambda + q/\Lambda} e^{-X/\Lambda} \left(1 - e^{-(\lambda + q/\Lambda)t} \right) + N_0(X, t) e^{-\lambda t} \quad (11.1-3)$$

Equation (11.1-3) shows that the result of erosion is equivalent to an increase in the decay constant or a reduced half-life. Correspondingly, accumulation leads to a shorter “effective” radioactive decay constant

$$\lambda_e = \lambda + q/\Lambda.$$

Equation (11.1-3) leads to two important results.

The Exposure Age. For simplicity, assume $N_0 = 0$. If the surface concentration is measured to be $C_S = N(X = 0, T)$ and the production rate is known from other considerations to be P , then substitution in (11.1-3) shows that the exposure age is

$$T = -\frac{1}{\lambda_e} \ln \left(1 - \frac{\lambda_e N_S}{P_0} \right) \quad (11.1-4)$$

(continued)

Saturation. For $t \gg 1/\lambda_e$, the surface concentration reaches an asymptotic value, the saturation concentration

$$N_S(t \rightarrow \infty) = \frac{P_0}{\lambda_e} \quad (11.1-5)$$

Now we have to distinguish between $q > 0$ (erosion) and $q < 0$ (accumulation).

$q > 0$: *erosion*

Consider the case where the surface is eroded at a constant rate of ε in $\text{g cm}^{-2} \text{s}^{-1}$. The erosion results in a loss of cosmogenic radionuclides similar to the loss by the radioactive decay, and the saturation concentration is $P/(\lambda + \varepsilon/\Lambda)$.

$q < 0$: *accumulation*

This case can be considered as negative erosion: $q = -\alpha$ in $\text{g cm}^{-2} \text{s}^{-1}$. We replace q with $-\alpha$, and assuming constant accumulation, we can replace t with $-X/\alpha$. The Eq. (11.1-3) then changes to a form that does not explicitly involve time, t , thus:

$$N(X) = \frac{P_0}{\lambda - \alpha/\Lambda} \left(e^{-X/\Lambda} - e^{-\lambda X/\alpha} \right) + N_0(X) e^{-\lambda X/\alpha} \quad (11.1-6)$$

As in the case of erosion, the second term describes the initial concentration of radionuclides in the surface layers (Lal 1988).

11.2 Terrestrial Solid Matter (Rocks, Ice)

The nature of the secondary particles has undergone considerable changes while travelling through the atmosphere. As Fig. 10.2.3-3 shows, the neutron intensity at sea level has been attenuated considerably. Furthermore, the nucleonic composition itself has changed; as Fig. 10.3-2 shows, the ratio of neutrons to protons increases from about 1 at the top of the atmosphere to about 5 at sea level. This illustrates that neutron production mechanisms dominate. Further, the mesons (Fig. 12.2.1-1) are less attenuated than the nucleonic component and they play an important role. In general, the several production rates are all very low, and require the use of very sensitive detection methods such as accelerator mass spectrometry (see Sect. 15.3).

Section 10.3 has shown that the cosmogenic production rate in the atmosphere decreases with a mean free path of about 160 g cm^{-2} corresponding to a decrease of a factor of about 650 ($e^{1033/160}$) from the top to the bottom of the atmosphere. Reference to Fig. 12.2.1-1 shows how the production rate continues to decrease underground for geomagnetic latitudes $>45^\circ$. The neutron component (evaporating and neutron disintegrations) continues to decrease at the same rate as above ground and quickly becomes negligible. Below $200\text{--}400 \text{ g cm}^{-2}$ underground the muonic

component becomes dominant because of its weaker interaction, resulting in a longer mean free path. In other words, in the atmosphere and the upper-most $\sim 500 \text{ g cm}^{-2}$ of the ground and ice the neutrons are dominant, while at greater depths the muons are more important. This will be discussed in detail in Sect. 12.2.1.

In the following, we will discuss in detail the production of ^{36}Cl in limestone and dolomite and ^{10}Be and ^{14}C in ice cores. The reason is that ^{36}Cl is quite unique in the sense that it can be applied in the study of nearly all types of rocks. Further it is highly soluble and therefore the danger of contamination by ^{36}Cl adsorbed to the surface of rocks is negligible. Last but not least the production of ^{36}Cl from Ca has three components: a spallation component, a fast muon component, and a muon capture component. This is an instructive example of the relative complexity of the cosmogenic production process.

^{10}Be and to a smaller extent ^{14}C are measured in ice cores for many purposes. In the following example, we therefore investigate whether in situ production of ^{10}Be and ^{14}C in ice has to be considered as an additional source to the production that originated in the atmosphere.

11.2.1 ^{36}Cl Production in Limestone and Dolomite

Limestone is a sedimentary rock composed mainly of calcium carbonate (CaCO_3) which makes up about 10% of all sedimentary rocks. Dolomite consists of calcium magnesium carbonate $\text{CaMg}(\text{CO}_3)_2$. In both minerals the interactions with Ca are the main source of ^{36}Cl , with a small but significant contribution from neutron activation of ^{35}Cl . We distinguish three production channels with Ca as the target element: spallation induced by fast neutrons, fast muons and slow muons. If ^{35}Cl is present, neutron activation also takes place through another three interactions; with epithermal and thermal neutrons slowed down from spallation events; from muon capture and from fast muon-induced reactions. Finally, ^{36}Cl will be produced by thermal neutrons originating from the U and Th content in rocks.

Spallation by fast neutrons leads to a ^{36}Cl concentration proportional to $e^{-X/\lambda}$ and contributes about 90% of the total production in Ca near the surface. Depending on the sample and the site, the geometry of the rock and shielding effects due to snow have to be taken into account.

Below 1 m capture of negative muons becomes the dominant reaction. The production rate at the depth z can be expressed as the product of the muon stopping rate R_{μ^-} (negative muons stopped in $1 \text{ g cm}^{-2} \text{ year}^{-1}$) times the yield of ^{36}Cl $Y(\mu^-)$ (^{36}Cl atoms per stopped negative muon):

$$P(\mu^-, {}^{40}\text{Ca})(z) = Y(\mu^-)R_{\mu^-}(z) \quad (11.2-1)$$

R_{μ^-} is 190 at sea level at high latitudes and decreases slowly with depth, yielding the production versus depth curve given in Fig. 12.2.1-1.

The fast muons do not contribute much and are not discussed here.

Whenever there is a significant amount of ^{35}Cl in the limestone (>20 ppm for 40 weight percent (wt%) Ca) or the dolomite (>10 ppm for 20 wt% Ca), ^{35}Cl will contribute at least 5% of the cosmogenic ^{36}Cl , primarily through capture of slow or epithermal neutrons. One source of the latter is due to the slowing down by elastic scattering of the fast neutrons in the nucleonic cascade (Fig. 10.2.1-1) (many interact with other atoms before they do so).

An important transition effect occurs in the slow neutron population at the Earth's surface. The abrupt change in the nature of the neutron absorbing elements that occurs going from air to soils and rocks results in a low-energy flux which is not in equilibrium with the flux of the high-energy neutrons. Since the production of thermal neutrons is higher in soils and rocks, diffusion into the air takes place:

$$F_{\text{sp,epi}} = k_{\text{epi}} e^{-z/\Lambda} - k_{\text{esc,epi}} e^{-z/L_{\text{esc,epi}}} \quad (11.2-2)$$

$$F_{\text{sp,th}} = k_{\text{th}} e^{-z/\Lambda} - k_{\text{esc,th}} e^{-z/L_{\text{esc,th}}} + k_{\text{corr}} e^{-z/L_{\text{esc,epi}}} \quad (11.2-3)$$

where $F_{\text{sp,epi}}$ is the spallation induced epithermal neutron flux ($\text{cm}^{-2} \text{ year}^{-1}$). The first terms describe the exponential decrease of the epithermal and thermal n -fluxes with depth z with a mean free path $\Lambda = 160 \text{ g cm}^{-2}$. The second terms are a correction to account for loss of neutrons to the atmosphere near the surface of the rock as discussed above. $k_{\text{esc,epi}}$ and L_{esc} depend upon the elemental composition of the rock. Finally k_{corr} is a correction for an increase of the thermal flux due to the thermalization of epithermal neutrons.

Figure 11.2.1-1 shows a comparison of calculated and observed epithermal and thermal neutron fluxes in a concrete block. The good agreement confirms that the basic processes are understood. It is important, however, to note that a changing water content affects the thermalization of neutrons and also has to be accounted for. Examination of the figure shows that the thermal neutron flux (and hence their

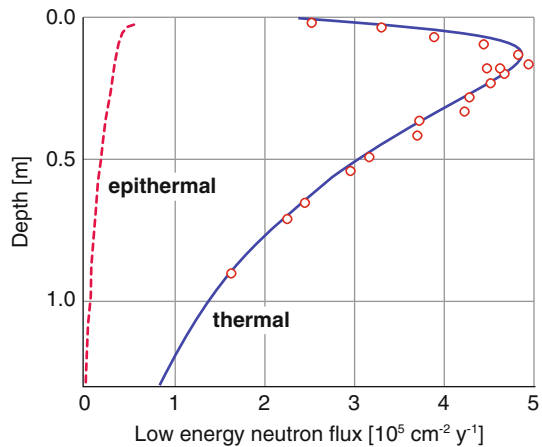


Fig. 11.2.1-1 Calculated epithermal and thermal neutron fluxes in a concrete block compared to measured thermal neutrons. After Alfimov and Ivy-Ochs (2009)

production of ^{36}Cl) rises rapidly in the first 15–20 cm, and then decreases rapidly with increasing depth.

Section 20.5 discusses hydrological applications of ^{36}Cl , and outlines experimental measurements that have been made in deep artesian basins.

In situ production of ^{10}Be , ^{26}Al , and ^{36}Cl in rocks has developed during the past decades into an important new field in the Earth sciences to determine exposure ages and erosion rates. For more detail, see Chaps. 21 and 23.

11.2.2 ^{10}Be and ^{14}C Production in Ice

Ice (and in particular polar ice) is one of the most important archives for ^{10}Be in order to reconstruct the history of solar activity. These reconstructions are based on the assumption that the ^{10}Be atoms found in an ice sample have originated in the atmosphere where they were produced by the interaction of cosmic rays with oxygen and nitrogen (Chaps. 10 and 13). From the discussions in this section it is clear that ^{10}Be is also produced in situ in the ice and we now consider whether this is a significant contribution compared to that from the atmosphere. For comparison, we also compute the ^{14}C production rates.

The in situ ^{10}Be and ^{14}C concentrations were calculated using (11.1-3) for a number of different altitudes and ice accumulation rates based on ^{10}Be production values given by Lal et al. (1987). Figure 11.2.2-1 shows the situation for an altitude of 3 km corresponding to drilling sites in central Greenland and Antarctica. The concentrations increase down to a depth of several hundred g cm^{-2} (corresponding to several meters) below which there is no significant production and a plateau is reached in the case of ^{10}Be . By contrast, ^{14}C starts to decrease below 10^3 to 10^4 g cm^{-2} due to the rather short ^{14}C half-life of 5,730 years.

Figure 11.2.2-1 shows that for an ice accumulation rate of 20 cm year^{-1} , typical for the GRIP and GISP2 ice cores in Greenland, the in situ ^{10}Be production is 600 atoms g^{-1} , compared with the measured average of $15,000 \text{ atoms g}^{-1}$. This implies an in situ contribution of about 4%. For Dome C in Antarctica with a considerably lower ice accumulation rate of 2.5 cm year^{-1} , the in situ production is about $5,000 \text{ atoms g}^{-1}$. Figure 21.3.2-1 shows that the measured ^{10}Be concentration at Dome C is about $50,000 \text{ atoms g}^{-1}$, indicating an in situ production of about 10%.

As discussed in Sect. 13.4, the tropospheric ^{10}Be that reaches the polar caps has originated in latitudes between 30° and 90° , and the degree of solar and geomagnetic modulation depends strongly on the latitude of origin. For example, Fig. 13.4-4 shows that only 19% of the tropospheric component originates in the polar cap, and consequently experiences the greatest level of solar modulation. The in situ production (occurring at very low cut-off rigidities) will also experience this highest level of modulation. This will not affect the interpretation of the ^{10}Be data greatly, except for those cases where the accumulation rate is low. The ^{10}Be in situ production rate drops rapidly going to lower altitudes, and will not be a problem

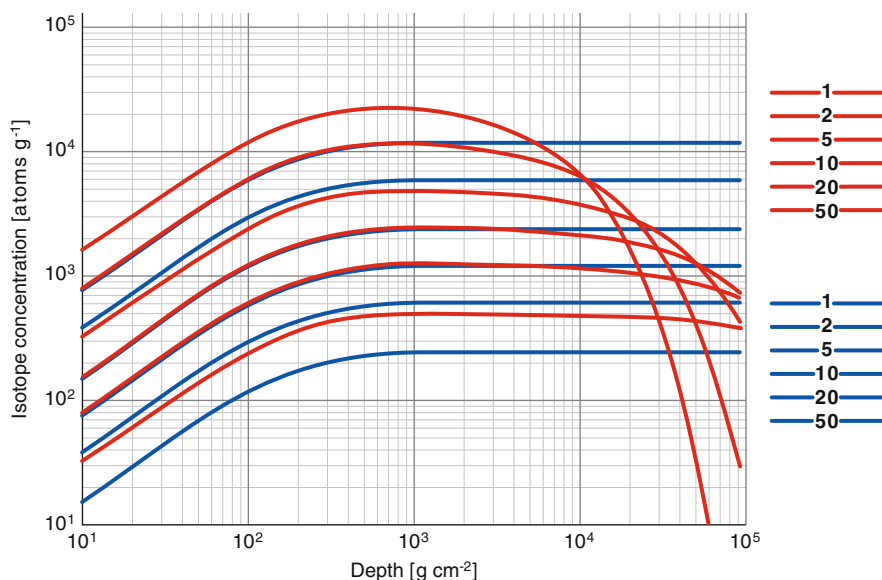


Fig. 11.2.2-1 ^{10}Be (blue lines) and ^{14}C (red lines) concentrations in polar ice due to in situ production for different ice accumulation rates ranging from 1 to 50 cm year^{-1} at the altitude of 3 km above sea level. After Lal et al. (1987)

under any circumstances. We note, however, that there may be an additional contribution from ^{10}Be transported by dust (see Sect. 21.3.5).

Measurements of in situ produced ^{14}C in polar ice have also been used to reconstruct the history of solar activity (Lal et al. 2005). The advantage of this approach is that in polar ice cores changes in the geomagnetic field intensity do not affect the production rate. The disadvantages are that the temporal resolution is low and that $^{14}\text{CO}_2$ in air bubbles occluded in ice can interfere with the in situ produced ^{14}C . In addition, in situ production in an ice core drilled for example in central Greenland can become a problem. The production rate of ^{14}C at an altitude of 3,000 m is 150 $\text{atoms g}^{-1} \text{ year}^{-1}$. According to Fig. 11.2.2-1 an ice core with an accumulation rate of 20 cm year^{-1} contains typically 1,000 ^{14}C atoms per gram of ice. Storing such an ice core at the surface for a year after drilling will lead to a “contamination” of 15%, which will become even larger if the ice is stored later in a freezer for several years before extracting the ^{14}C .

11.3 Extraterrestrial Solid Matter

For the galactic cosmic radiation (GCR) with $E > 1 \text{ GeV}$ (see Fig. 5.7.3-1), the interactions with extraterrestrial solid matter such as lunar and meteoritic materials and the subsequent production of cosmogenic radionuclides are very

similar to those processes in the atmosphere. Thus the cosmic ray flux is attenuated with depth with a mean free path $\lambda \approx 160 \text{ g cm}^{-2}$, and secondary particles are produced. They, in turn, initiate spallation reactions that generate the cosmogenic radionuclides.

The solar cosmic radiation (SCR, Chap. 8), however, behaves quite differently. Its spectrum is much softer, with the majority of the particles having $E < 100 \text{ MeV}$ (Fig. 8.3.1-2). The upper panel in Fig. 10.2.2-1 shows that they will lose their energy through ionization very quickly; the bottom panel shows that they will come to the end of their range (i.e. stop) within the first 10 g cm^{-2} (Note that while Fig. 10.2.2-1 is for air, the same principle applies for meteoritic and lunar materials). That means the galactic cosmic rays generate cosmogenic material deep into the target, while the SCR only produce a “skin” on the exposed surfaces. We will return to this later.

Section 11.1 pointed out three main differences of detail (spectrum, different target elements, and immobility of the cosmogenic radionuclides) between the cosmogenic interactions in terrestrial matter and the atmosphere and these apply to extraterrestrial material as well. With respect to the first, meteorites are moving targets. Some are in orbits close to that of Earth and others are in highly elliptical orbits whose aphelions may be $>20 \text{ AU}$ from the Sun. Kepler’s second law of planetary motion indicates that meteorites in highly eccentric orbits spend the majority of the time near aphelion (i.e. a long way from the Sun). The cosmic ray flux in the heliosphere depends strongly on distance from the Sun (Sect. 5.7) and consequently the exposures will depend strongly on the nature of the orbit. Some meteorites will primarily sample the GCR far out in the heliosphere and others close to the orbit of Earth. Further, the degree of solar modulation (characterized by the modulation function, Φ) varies strongly with time (Sects. 5.7, 7.2 and 7.3). Thus two long-period (hundreds of years, say) meteorites on exactly the same orbit, but out of phase with one another, may experience very different exposures depending on their phasing with respect to the Grand Minima in solar activity. The last of the three properties (immobility) permits study of the depth profiles of the cosmogenic radionuclides and permits production models to be tested.

Figure 11.3-1 is an excellent demonstration of the differences between the effects of GCR and SCR on extraterrestrial matter. It shows that the depth profiles of ^{10}Be , ^{26}Al , and ^{36}Cl measured in lunar rock 64455 (collected by the crew of Apollo 16) agree well with the calculated values. As discussed above, it is necessary to take into account both the galactic (GCR) and the solar (SCR) components of the cosmic radiation in the uppermost 10 g cm^{-2} . The differential flux of the SCR component can be described by:

$$\frac{dJ}{dR} = k \cdot e^{-R/R_0} \quad (11.3-1)$$

with J the differential proton flux as a function of the rigidity R in MV (Rigidity is defined in Sect. 5.3). The rigidity R_0 defines the slope of the spectrum and ranges from 20 to 200 MV depending upon individual flares. We note in passing that the

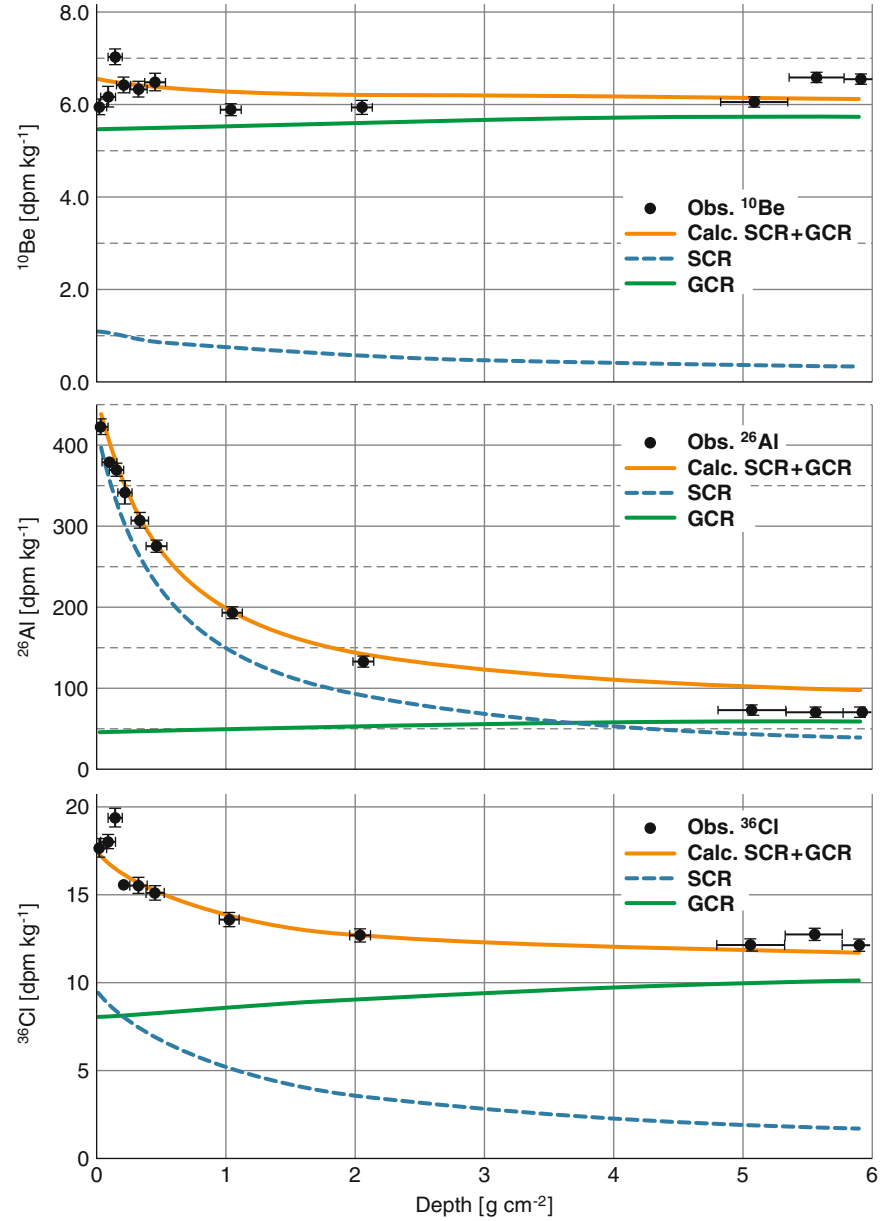


Fig. 11.3-1 Comparison of calculated and measured depth profiles of ^{10}Be , ^{26}Al , and ^{36}Cl in lunar rock 64455 collected by the crew of Apollo 16. The depth profiles reflect the sum of the galactic cosmic ray component (*green line*) which changes only slightly within the top 6 g cm^{-2} , and the solar cosmic ray component (*blue dashed line*) which is less energetic and therefore shows a clear decrease with increasing depth [after (Nishiizumi et al. 2009)]. The concentrations of the cosmogenic radionuclides are given in dpm kg^{-1} (decays per minute per kg of rock)

spectra of SCR are sometimes expressed as exponential functions as here, and sometimes as power laws (Figs. 8.3.1-1 and 8.3.1-2).

Small values of R_0 in (11.3-1) mean that the solar spectrum decreases rapidly with increasing rigidity, meaning that most solar particles are of relatively low energy and cannot penetrate deeply into the rock (as discussed above). All production rates therefore show a decreasing trend with increasing depth for the SCR component (blue lines). In contrast, the GCR component shows a slightly increasing trend with depth (green line) which is due to the generation of secondaries which will reach a maximum between 100 and 200 g cm⁻² as in the atmosphere.

In the case of the SCR, the strong differences between the decreasing trends of the three cosmogenic radionuclides are caused by the characteristics of their excitation functions (see cross sections, Sect. 10.3.2). While the thresholds for the production of ¹⁰Be by protons are relatively high (around 100 MeV), they are relatively low for ²⁶Al produced from Mg and Si. As a result, the production of ²⁶Al in the upper ~4 g cm⁻² is dominated by the SCR, while in the case of ¹⁰Be the SCR component only contributes some 10% to the total production rate. In the case of ³⁶Cl both components are about equal in the surface layer.

Analysis of the depth profiles yields an erosion rate of less than 0.5 mm per million years. This erosion rate for the ²⁶Al profile is consistent with a mean spectral shape given by $R_0 = 90$ MV, and an omni-directional proton flux of 196 cm⁻² s⁻¹ for energies above 10 MeV.

Overall the good agreement between the measured and the calculated depth profiles indicates that the production models are appropriate to describe the production rates of cosmogenic radionuclides in extraterrestrial matter.

References

- Alfimov V, Ivy-Ochs S (2009) How well do we understand production of Cl-36 in limestone and dolomite? *Quat Geochronol* 4(6):462–474
- Lal D (1988) In situ-produced cosmogenic isotopes in terrestrial rocks. *Annu Rev Earth Planet Sci Lett* 16:355–388
- Lal D, Nishiizumi K, Arnold JR (1987) Insitu cosmogenic H-3, C-14, and Be-10 for determining the net accumulation and ablation rates of ice sheets. *J Geophys Res-Solid Earth Planet* 92(B6):4947–4952
- Lal D, Jull AJT, Pollard D, Vacher L (2005) Evidence for large century time-scale changes in solar activity in the past 32 Kyr, based on in-situ cosmogenic C-14 in ice at Summit, Greenland. *Earth Planet Sci Lett* 234(3–4):335–349
- Nishiizumi K, Arnold JR, Kohl CP, Caffee MW, Masarik J, Reedy RC (2009) Solar cosmic ray records in lunar rock 64455. *Geochim Cosmochim Acta* 73(7):2163–2176

Chapter 12

Alternative Production Mechanisms

12.1 Introduction

In Chapter 10 we described the production of radionuclides by cosmic rays in the atmosphere. We have seen that the limited number of target elements in the atmosphere (mainly N, O, and Ar) means that only a small number of different cosmogenic radionuclides are produced. In Sect. 11 we have then discussed the production rate of cosmogenic radionuclides in other environmental systems such as rocks, meteorites, and water. In this section we provide an overview of other production mechanisms which may be important for specific radionuclides in general or for common radionuclides at specific sites. First in Sect. 12.2 we consider natural production mechanisms related to cosmic rays and to radioactive decay underground. Then in Sect. 12.3 we address the production of radionuclides related to human activities. Among those are the radionuclides produced in nuclear bomb explosions and power plants. In recent times there is also a growing number of radionuclides mainly produced by accelerators for applications in industry, research, and medicine.

12.2 Natural Production Mechanisms

12.2.1 *Cosmic Ray Induced Reactions*

The flux of secondary particles (mainly neutrons and protons) in the atmosphere is strongly attenuated as the nucleonic cascade descends (see Sect. 10.2-1). The flux is very low at sea level ($\sim 1,000 \text{ g cm}^{-2}$), whereas on the top of high mountains the fluxes are up to a factor of ten higher (Fig. 10.2.3-3). The top-most 4 m of soil and rock then has a similar shielding effect as the whole atmosphere above it (for an average density of 2.7 g cm^{-3} ; $400 \text{ cm} \times 2.7 \text{ g cm}^3 = 1,080 \text{ g cm}^{-2}$). As a consequence the neutrons and protons generated in the atmosphere do not penetrate

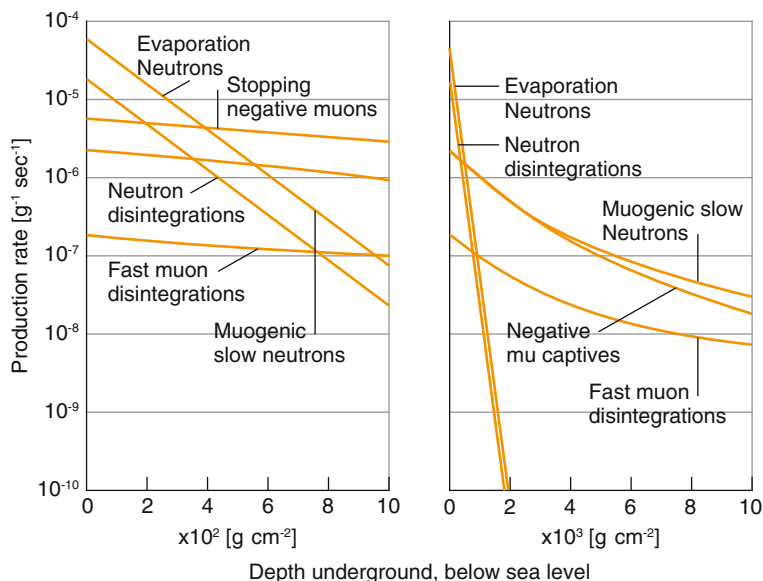


Fig. 12.2.1-1 Contribution of different particles to the production of radionuclides below ground for geomagnetic latitudes $>45^\circ$. After Lal (1987)

more than approximately a metre into the Earth's crust. However, there are other mechanisms that initiate nuclear reactions (although with generally lower production rates) that become the dominant sources of cosmogenic radionuclides.

Box 12.2.1.1 provides background information on the elementary particles of interest.

Figure 12.2.1-1 summarizes the cosmogenic production process that occurs underground. As outlined above, neutron-induced processes cease to be the most important source at a depth of $\sim 4 \times 10^2 \text{ g cm}^{-2}$; that is $\sim 1.5 \text{ m}$ underground. They become trivial below a depth of 2 m. Cosmogenic production below that depth is dominated by interactions initiated by the muon components of the nucleonic cascade that penetrate to relatively great depths by virtue of their long mean free path ($\sim 250 \text{ g cm}^{-2}$). The production of ^{36}Cl by muons in limestone has been discussed in Sect. 11.2.1, and its applications in hydrology are discussed in Sect. 20.5..

12.2.1.1 BOX Muons and Neutrinos

Muons

Muons are electrically charged elementary particles. Together with the electrons, the taus, and the neutrinos, they form the family of the so-called leptons with spin $\frac{1}{2}$. The mass of the muon is 105.7 MeVc^{-2} , which is equivalent to a mass 207 times that of an electron. As a first approximation,

the interaction of a muon with matter can be considered as that of a heavy electron. However, as a result of the larger mass, the interaction (mainly the production of Bremsstrahlung) is much weaker and therefore they have a much longer range. Having slowed down, a negative muon can be captured by a proton, usually resulting in the production of a neutron and a neutrino. If the proton is part of a heavy nucleus other particles (in most cases a neutron, but sometimes charged particles and gamma rays) can then be emitted.

Muons are unstable with a mean lifetime of 2.2 μs . A muon decays into an electron and two neutrinos:

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$$

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$$

The common negative muon decays into an electron, an electron–anti-neutrino, and a muon–neutrino. The less common positive muon decays into a positron, an electron–neutrino, and a muon–antineutrino.

As Fig. 10.2.1-1 illustrates, muons originate in cosmic ray interactions in the upper atmosphere. High-energy primary particles create pions which decay after a mean lifetime of 2.6×10^{-8} s into a muon and neutrinos. Most of these reactions take place between 10 and 20 km above sea level and about 100 muons reach every square metre of the earth's surface per second. Non-relativistic considerations would say that a muon travelling at close to the speed of light would only travel 660 m (2.2×10^{-6} s $\times$ 3×10^8 μs^{-1}) in their average life-time; the fact that they reach surface is a consequence of the “dilation of time” in the theory of relativity (i.e. “their clocks run slow” – see Box 5.3.1).

Neutrinos

The existence of neutrinos was postulated by Pauli in 1930 in order to preserve the laws of conservation of energy, momentum, and angular momentum in beta decay when a neutron turns into a proton and an electron:

$$n \rightarrow p^+ + e^- + \bar{\nu}_e$$

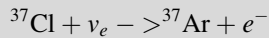
It was more than 20 years before the neutrino was detected experimentally, near a nuclear reactor in which there was a very high rate of beta decays. The reason is that neutrinos interact very weakly with matter, they pass right through it, and they are therefore extremely difficult to detect.

(continued)

There are two ways neutrinos do interact with matter:

1. In the “charged current interaction”, a high-energy neutrino turns into a lepton (electron, muon, and tauon).
2. In the “neutral current interaction” the neutrino transfers some of its energy and momentum to a target particle. Provided the target particle is sufficiently light and charged (as for example an electron) it may be accelerated sufficiently to cause a nuclear reaction or to produce Cherenkov radiation.

Both interactions have been used to detect neutrinos from the Sun and supernovae. To reduce background problems induced by cosmic rays, neutrino detectors are built deep underground. The first experiment took place in the former Homestake mine in South Dakota. Almost 470 tons of tetrachloroethylene were used as target material in a big tank (Fig. 12.2.1-2) (Davis 1994). Every day a few atoms of ^{37}Ar were produced by the reaction:



The ^{37}Ar was then extracted from the huge tank and transferred to a proportional counter that detected the electron which ^{37}Ar releases when it decays back into ^{37}Cl with a half-life of 35 days.



Fig. 12.2.1-2 Neutrino experiment in the Homestake Gold mine in South Dakota about a mile under solid rocks (Davis 1994)

12.2.2 Radioactive Decay-Induced Reactions

The solar system is composed of elements which were produced during nucleosynthesis in stars and subsequent supernova explosions. Fusion reactions in stars lead to the formation of the elements up to iron while the high neutron fluxes in supernova explosions resulted in the production of the heavier elements up to uranium. Many of the radionuclides so produced decayed quickly; however, those with half-lives of $>10^8$ years are still present to some extent on Earth and are called the primordial radionuclides. There are about 20 primordial radionuclides, the most prominent being ^{238}U , ^{235}U , ^{232}Th , and ^{40}K . Because they were part of the planetary nebula from which the solar system was formed, they are found everywhere on Earth, but with concentrations that vary as the consequence of geochemical processes.

To be relevant to the production of radionuclides, the primordial nuclides must either turn into another radionuclide by radioactive decay or fission, or emit an alpha particle or a neutron which then interacts with matter producing a new radionuclide. ^{40}K does not provide any new radionuclides because it decays with a probability of 89% into ^{40}Ca and a probability of 11% into ^{40}Ar , both being stable. Natural uranium consists of 99.3% of ^{238}U , 0.7% of ^{235}U and a negligible amount (0.006%) of ^{234}U . It has an average concentration of ~ 3 ppm (parts per million) in the Earth's crust, and ~ 3 ppb (parts per billion) in the oceans. Ores contain between 1 and 23% of uranium oxides. We now examine the role of uranium in the production of the radionuclides that augment those produced by the cosmic radiation.

Figures 12.2.2-1 and 12.2.2-2 show the decay chains of ^{238}U and ^{235}U . For each daughter of the decay chain the atomic mass number, the atomic number, the half-life, and the released energy are given. All the nuclides decay by either β^- or α emission. Sometimes the configurations of neutrons and protons are such that both decays are possible leading to different decay pathways. However, ultimately all decay chains end at the stable isotopes ^{206}Pb and ^{207}Pb , respectively.

The α -particles emitted by the daughter nuclides of ^{238}U and ^{232}Th produce neutrons by interacting with the surrounding elements: Si, O, Al, Mg, and others. These neutrons can then produce radionuclides such as ^{36}Cl in exactly the same way that neutrons do in the atmosphere (Sect. 10.3). Trace elements such as B, Gd, Sm, and Li in the rock will act as neutron absorbers, thereby reducing the neutron flux. The neutron flux therefore depends on the U and Th content, as well as the elemental composition of the rock itself. The mean free path for thermal neutrons is approximately 50 cm so the production process can vary strongly over spatial scales of ~ 1 m. Estimated neutron production rates for different rock types are (Andrews et al. 1986):

Granite: $P = 2.04[\text{U}] + 0.70[\text{Th}]; \sigma_m = 0.0125$

Sandstone: $P = 1.31[\text{U}] + 0.38[\text{Th}]; \sigma_m = 0.0123$

Limestone: $P = 1.11[\text{U}] + 0.29[\text{Th}]; \sigma_m = 0.0055$

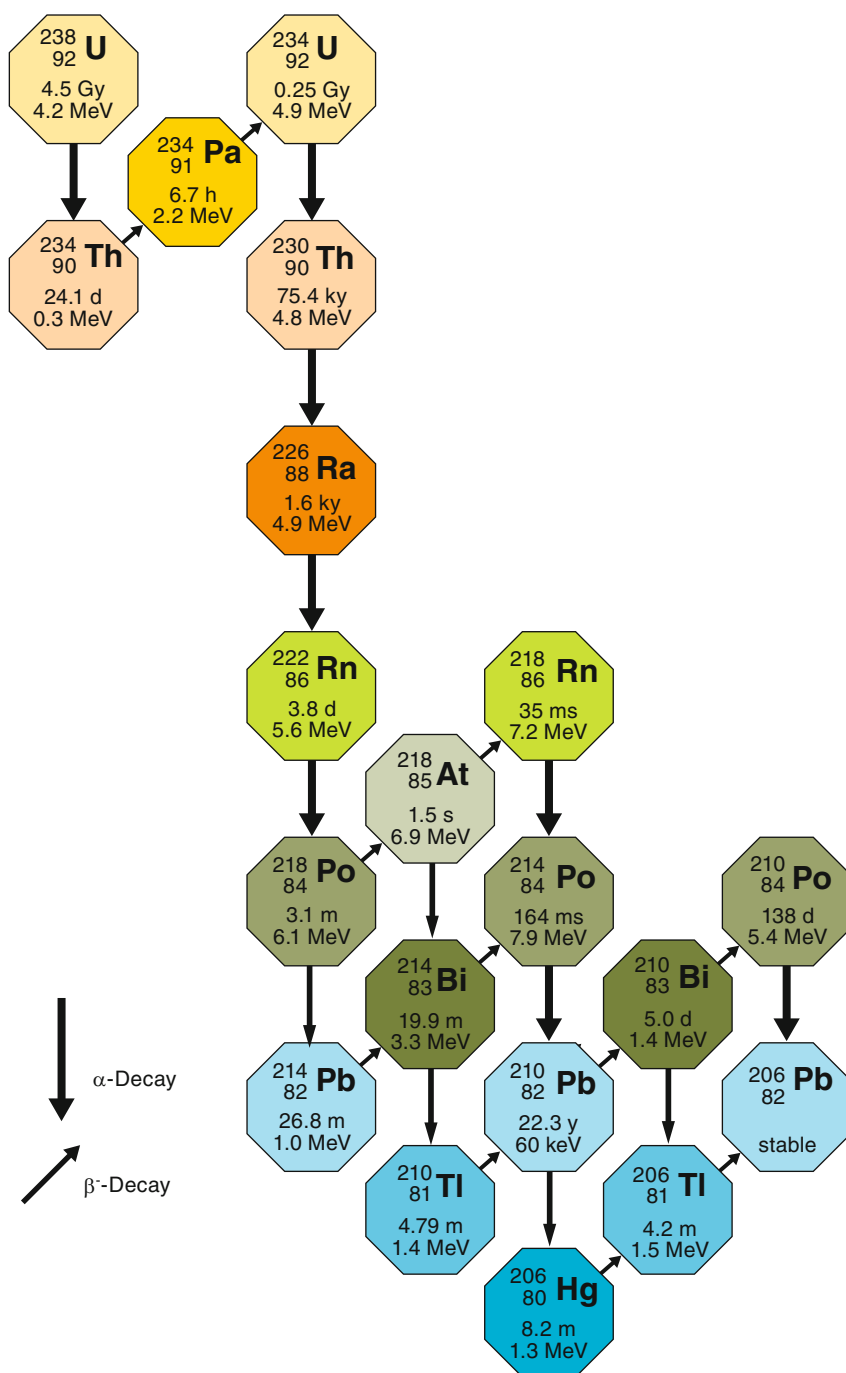


Fig. 12.2.2-1 Decay chain of ^{238}U leading to ^{206}Pb . For each daughter of ^{238}U the atomic mass number, the number of protons, the half-life, and the released energy are given. In cases when different α - and β^- decays can occur the less probable decays are indicated with a *thinner arrow* and the energy corresponds to the more probable decay

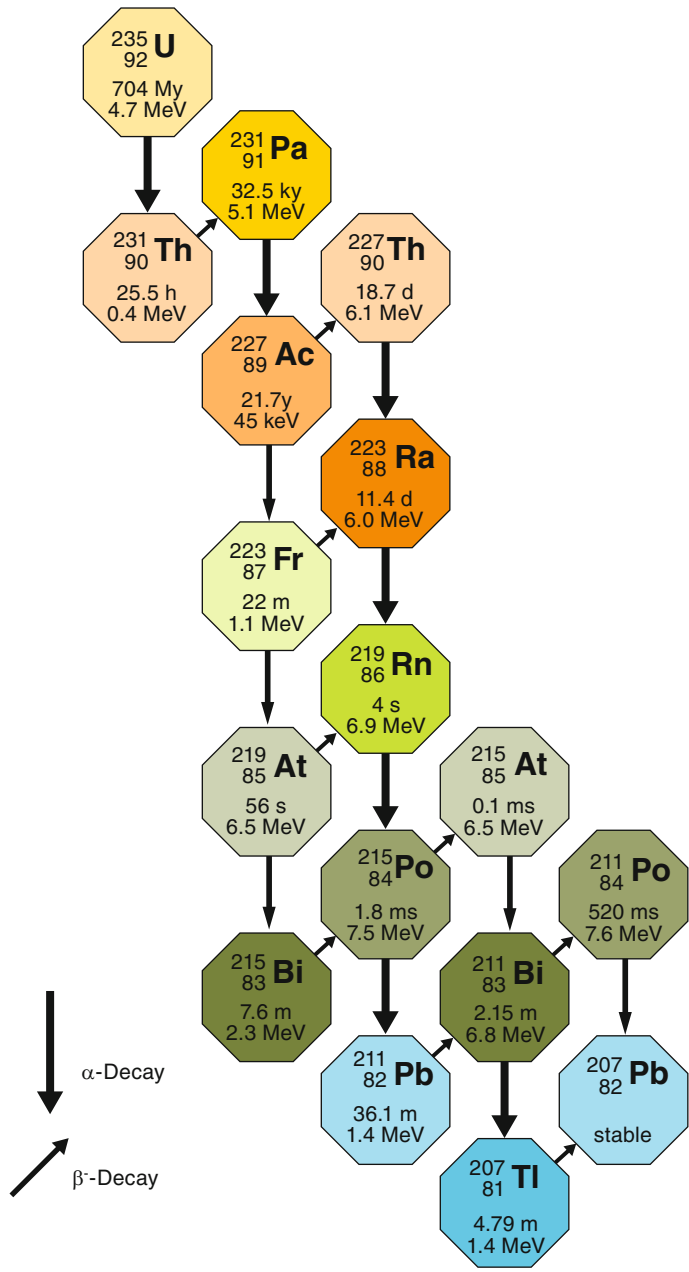


Fig. 12.2.2-2 Decay chain of ^{235}U leading to ^{207}Pb . See the caption of Fig. 12.2.2-1 for further details

with P the production in neutrons per gram of rock per year: $[U]$ and $[Th]$ are the concentrations in ppm, and σ_m is the weighted total neutron absorption cross section in $\text{mol} \times \text{barn}$ per gram.

Combining these values provides the thermal neutron flux F_n in neutrons $\text{cm}^{-2} \text{s}^{-1}$:

$$F_n = 10^{-5}(\alpha[U] + \beta[Th])$$

The α and β values depend on the rock composition and are estimated to be (Andrews et al. 1986):

Granite: $\alpha = 0.86$ $\beta = 0.30$

Sandstone: $\alpha = 0.56$ $\beta = 0.10$

Limestone: $\alpha = 1.07$ $\beta = 0.28$

Fission is another source of neutrons of relevance. Heavy nuclei such as uranium and thorium undergo fission in which the nucleus splits into smaller parts, often accompanied by the emission of neutrons and gamma rays. Fission can occur spontaneously or be induced by an incident neutron. Some of the fission products are themselves radioactive. Well-known examples of these are ^{137}Cs , ^{90}Sr , and ^{129}I . The neutrons emitted in a fission reaction can induce new fissions or interact with other atoms producing radionuclides as outlined in the previous paragraphs.

12.3 Anthropogenic Production Mechanisms

12.3.1 Nuclear Power Plant and Nuclear Bomb-Induced Reactions

The main difference between a nuclear power plant and a nuclear fission bomb (atomic bomb) is that the fission rate in a power plant is kept stable, and great care is taken to minimize the escape of neutrons and radionuclides. In a nuclear bomb, on the other hand, the neutron flux is unconstrained and great quantities of radionuclides escape into the atmosphere. The amount of energy released in a bomb in less than a second is in the range of that released by 1–500 kilotons of TNT (Trinitrotoluol). Both nuclear bombs and power plants are sources of essentially the same radionuclides, through two distinct mechanisms. The first source is from fission reactions.

As shown in Fig. 12.3.1-1, the fission products form a bimodal distribution with masses in the range from 80–110 and 130–150 leading to a sum of 235–239 depending on the nuclide which underwent fission. Typical representatives of the nuclides in the lighter peak are ^{90}Sr ($T_{1/2} = 28.9 \text{ y}$), ^{85}Kr ($T_{1/2} = 10.76 \text{ y}$), ^{79}Se ($T_{1/2} = 295 \text{ ky}$), ^{93}Zr ($T_{1/2} = 1.53 \text{ My}$), and ^{99}Tc ($T_{1/2} = 211 \text{ ky}$). For the

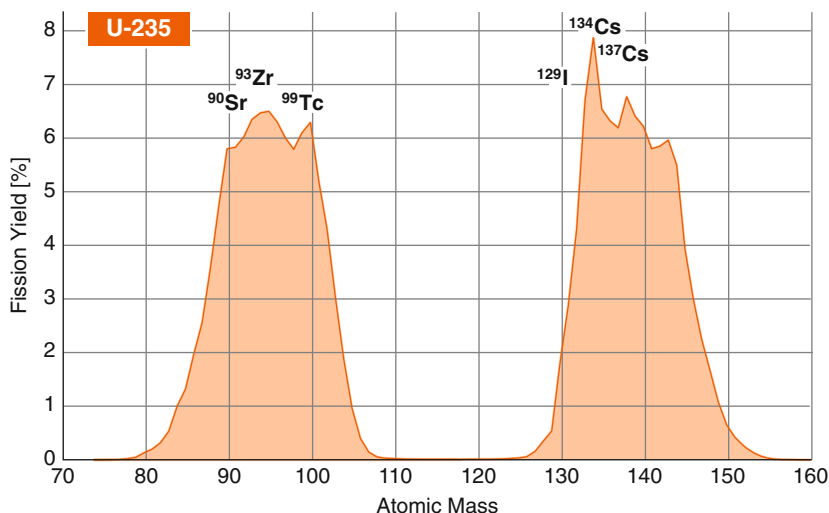


Fig. 12.3.1-1 Fission product yields in percent as a function of mass for thermal neutron induced fission of ^{235}U (source: http://en.wikipedia.org/wiki/Nuclear_fission_product)

heavier peak examples are ^{131}I ($T_{1/2} = 8.1$ d), ^{129}I ($T_{1/2} = 15.7$ My), ^{134}Cs ($T_{1/2} = 2.06$ y), ^{137}Cs ($T_{1/2} = 30.23$ y), and ^{126}Sn ($T_{1/2} = 230$ ky).

These fission products remain in the fuel rods of the power plants and are then extracted in specialized reprocessing plants. Most of the nuclides are prepared for storage in nuclear waste repositories.

In the case of the bombs, however, all these fission products are injected into the atmosphere. As a result of the strong thermal uplift a large fraction of the nuclides enters the stratosphere and is then distributed globally in a manner similar to the cosmogenic radionuclides that are produced in the stratosphere (Chapter 13).

The second source of radionuclides produced by nuclear power plants and bombs is due to neutrons interacting with the surrounding atoms. In the case of due to nuclear power plants great effort is to keep the production of radionuclides by neutron activation as small as possible and to reduce their release into the environment. In the case of nuclear bombs neutron activation is an important source with world-wide implications. Examples of the radionuclides produced in nuclear reactors are ^{14}C , ^{54}Mn , ^{58}Co , ^{60}Co , and ^{65}Zn . Good examples of bomb-induced radionuclides are ^3H and ^{36}Cl . Both are produced by neutron reactions: $^{14}\text{N}(n,^3\text{H})^{12}\text{C}$ and $^{35}\text{Cl}(n,\gamma)^{36}\text{Cl}$. Sea salt is the primary source of chlorine for the second reaction, and therefore the majority of the ^{36}Cl was produced between 1946 and 1960, when a number of bomb tests were carried out on atolls and ships (Sect. 19.3).

In Chap. 19 the ^{36}Cl fallout measured in ice cores is compared to model calculation of the atmospheric transport and deposition. ^3H and ^{14}C , which are

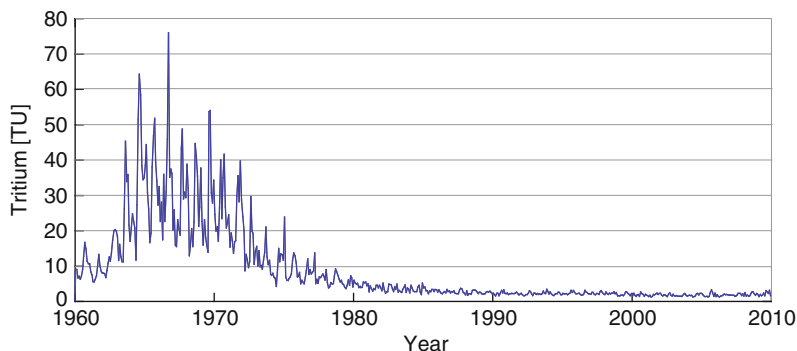


Fig. 12.3.1-2 Tritium concentrations in TU (Tritium Units) in monthly rain water samples from Kaitoke, New Zealand. After Morgenstern and Taylor (2009)

both produced by neutron reactions with nitrogen, peaked in 1963 just before the implementation of the nuclear test ban (Sect. 19.3).

Figure 12.3.1-2 shows the ^3H in monthly rain water samples from Kaitoke, New Zealand (Morgenstern and Taylor 2009).

Note the strong annual cycle due to transport and deposition processes. Since the majority of nuclear tests took place in the northern hemisphere, the annual mean ^3H concentrations are a factor of about 50 greater than those in the southern hemisphere.

12.3.2 Research, Industrial, and Medical Induced Reactions

From the beginning of the nuclear age, experimental nuclear physics has depended on the study of artificially induced nuclear reactions. Ever-larger accelerators have been built to study very high-energy interactions. At present, the highest energies are reached by the Large Hadron Collider (LHC) at CERN. Two beams of hadrons (protons or lead ions) are accelerated up to 7 TeV (7,000 GeV) in opposite directions in a circular tube of 27 km length before they collide and produce a variety of secondary particles, in exactly the same way that high energy cosmic ray particles interact with the atmosphere and the Earth's crust. It is therefore obvious that all accelerators which generate ions of high enough energies to induce nuclear reactions will contribute to the global production of radionuclides (Simakov et al. 2008). However, their production is relatively small compared with natural production and therefore is not discussed in detail.

Today, there are many medical and industrial applications which require radionuclides with specific properties. An important criterion in selecting a specific radionuclide must always be the avoidance of contamination of the environment, and minimisation of the exposure of people. An example of a common application

is the measurement of the thickness of a product such as paper or metal foil by continuously measuring the absorption of a gamma ray beam emitted from a point source and crossing the target material.

In medical sciences radionuclides are used for diagnostic and therapeutical purposes. In all medical applications it is especially important to expose patients to low doses of radioactivity. This can be achieved by using radionuclides with short radioactive and biological half-lives which concentrate in the area of interest. In therapeutical applications the energy released by the decay of a radionuclide is used to destroy carcinogenic cells.

References

- Andrews JN, Fontes J-C, Michelot J-L, Elmore D (1986) In-situ neutron flux, ^{36}Cl production and ground water evolution in crystalline rocks at Stripa, Sweden. *Earth Planet Sci Lett* 77:49–58
- Davis R (1994) A review of the homestake solar-neutrino experiment. *Prog Part Nucl Phys* 32(32):13–32
- Lal D (1987) Cosmogenic nuclides produced in situ in terrestrial solids. *Nucl Instrum Methods Phys Res B* 29:238–245
- Morgenstern U, Taylor CB (2009) Ultra low-level tritium measurement using electrolytic enrichment and LSC. *Isotopes Environ Health Stud* 45(2):96–117
- Simakov SP et al (2008) Activation of the IFMIF prototype accelerator and beam dump by deuterons and protons. *Fusion Eng Des* 83(10–12):1543–1547

Chapter 13

Transport and Deposition

13.1 Introduction

Having successfully propagated through the heliosphere and the geomagnetic field (Sects. 5.7 and 5.8), a cosmic ray enters the atmosphere. In Chap. 10, we have described how a cascade of secondary particles (protons and neutrons) then develops, which attenuates with atmospheric depth. Due to the geomagnetic shielding which excludes primary particles below the cut-off rigidity, the cascade is both depth and latitude dependent. The cascade interacts with atmospheric atoms to generate the cosmogenic radionuclides. As a result of all these processes, the production rate is proportional to the proton and neutron fluxes and therefore depends strongly on atmospheric depth, geomagnetic latitude, solar activity, and geomagnetic field intensity, as summarized by the figures in Chap. 10.

One of the main fields of applications of cosmogenic radionuclides is the reconstruction of the past solar activity and its potential effects on the Earth and the climate system. An ideal way to track changes in the production rate of the cosmogenic radionuclides is, therefore, to stay at one specific site in the atmosphere, preferentially at a high altitude and a high latitude where the production rate is large. That is exactly what neutron monitors have done since 1951 (see Sect. 6.3). However, if we want to go back thousands of years in time, we need an archive which preserves the cosmogenic radionuclides in a chronological sequence. As will be discussed in Chap. 14, ice and sediment cores are good examples of such archives for ^{10}Be and ^{26}Al , and tree rings are for ^{14}C .

As was shown in Chap. 10, most of the cosmogenic radionuclides are produced high up in the atmosphere, and consequently, the atoms must travel a long distance before they are stored in the archive. This results in two distinct interpretation problems:

- (a) The extent to which the archive contains nuclides originating at different geomagnetic latitudes for which the modulation effects are substantially different. For example, Fig. 10.3.3.2 shows that for a change in the modulation function from 200 to 400 MeV, the ^{10}Be production rate changes by 20% at

high latitudes, and $\sim 2\%$ at low latitudes. Clearly, the observed change in the archive will be strongly influenced by where the ^{10}Be has originated. Without knowing what mixing has occurred in the atmosphere, it will be impossible to deduce the change in the modulation function corresponding to an observed change in the data in the cosmogenic archive.

- (b) The global atmospheric circulation varies with the seasons, and from year to year. Different processes transfer the nuclides from the atmosphere into the archive; we refer to these as scavenging processes. So even if the production rate is constant, the flux of cosmogenic radionuclides reaching the archive will vary due to changes in the atmospheric circulation, and variations in the scavenging processes. Clearly, any changes introduced by variable atmospheric processes will interfere with any study of the manner in which the production rate has varied over time. These time-dependent atmospheric effects are sometimes referred to as “system effects” to distinguish them from “production changes”.

Figure 13.1-1 illustrates the nature of the interpretation problems introduced by (a). It plots the fractional decrease in the flux of ^{10}Be to a polar archive, for six hypothetical atmospheric mixing models, as a function of the modulation function, Φ . The extreme models, M1 (only high latitude production) and M6 (total global mixing), have both been used in the literature to interpret ^{10}Be observations from Greenland and

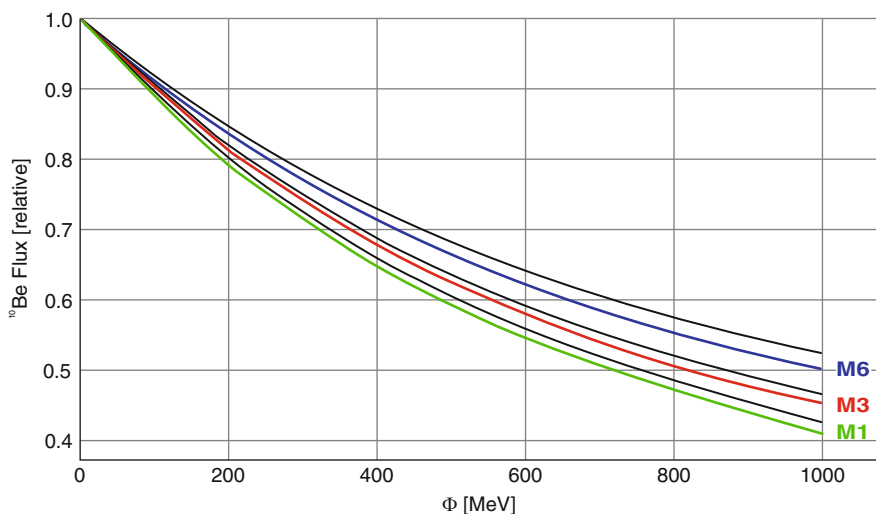


Fig. 13.1-1 Illustrating the manner in which the amplitude of the modulation of the ^{10}Be flux observed in polar regions depends on the nature of the inter-latitudinal mixing in the atmosphere. Model 6 (M6) is total global averaging, and this results in the smallest change in flux of ^{10}Be into the polar archive (compared to the LIS value of 100%) for a given degree of solar modulation. M1 is the case where the only ^{10}Be reaching the archive has been produced in the polar cap, resulting in the greatest change. M3 is a hypothetical model where production only from latitudes $>20^\circ$ has contributed to the archive. The ^{10}Be flux is relative to that reaching the polar archive if the local interstellar cosmic ray spectrum (the LIS) were incident on Earth (i.e. zero modulation). After McCracken (2004)

Antarctica. Consider now that the observed ^{10}Be observation is 0.6 of that without any modulation. The M1 curve would interpret this as implying $\Phi = 450$ MeV, while M6 implies $\Phi = 740$ MeV. In some analyses, the use of the global average (M6) has resulted in the unrealistic (and meaningless) conclusion that Φ attained negative values during the Maunder Minimum, while for the same observations, M1 resulted in positive values. These two mixing models are the limiting cases, and the mixing due to the real atmosphere will lie somewhere in between.

In summary, the atmosphere imposes averaging and filtering processes that change the amplitude of the “production” signal in the archive, and also adds noise to it due to system changes. The difficulty with the atmospheric averaging and filtering processes is that their properties are not well known, and they change from place to place and almost certainly over time – say between the Maunder Minimum and now. The purpose of this chapter is to provide some basic information on the various atmospheric processes, in order that we can better understand the general nature of the system effects, and to assist us in the interpretation of the data obtained from the archives.

It is important to note that the treatment of the atmospheric effects upon the ^{10}Be production signal, in particular, was rudimentary in the extreme until about 2005. It is reasonable to expect that our understanding will continue to improve in the near future using the mathematical modelling techniques discussed in Sect. 13.4. Reference will also be made to Chap. 19, where we discuss applications of the cosmogenic radionuclides to the study of the atmosphere itself.

The atmospheric transport and deposition processes are very important because they apply to all the cosmogenic radionuclides produced in the atmosphere. About 1/3 of them are deposited on the continents and are stored in ice sheets and soils. This reduces their mobility considerably. Transport effects also depend strongly on the geochemical properties of the nuclides. Insoluble ones such as ^{10}Be and ^{26}Al become attached to soil particles, while soluble ones such as ^{36}Cl and ^{129}I enter the water cycle (for details of the individual geochemical cycles, see Sect. 13.5). About 2/3 of the cosmogenic radionuclides are removed from the atmosphere into the ocean and continue their travel depending on their chemical properties. Chemically reactive nuclides such as ^{10}Be and ^{26}Al attach themselves to particles and are scavenged within 500–1,000 years into the sea-floor sediment (see Sect. 14.7.2).

13.2 Basics of the Atmosphere

As already mentioned (Box 10.1.1), the atmosphere is divided into several regions (troposphere, stratosphere, mesosphere, thermosphere, and exosphere) ranging from the Earth’s surface to about 10,000 km altitude and above. Since the pressure decreases approximately exponentially with altitude, the amount of atmospheric mass above 30 km is practically negligible ($<1.2\% = 12 \text{ g cm}^{-2}$). As we have seen in Chap. 10, there is little production of cosmogenic radionuclides above this atmospheric depth, and consequently, we only consider the stratosphere and the

troposphere in the following. The physical properties and mass motions within these two regions are quite different: there is a rather sharp boundary between them, and we now discuss the influence that each of these factors has on the cosmogenic radionuclides that reach an archive of interest.

Transport Processes and the Troposphere. The incoming electromagnetic solar radiation is not equally distributed over all latitudes. In the tropics, a daily average of about 480 W m^{-2} (for an albedo of 0.3) is absorbed by the Earth's surface, and this drops to below 100 W m^{-2} in Polar Regions. This results in a strong latitudinal temperature gradient, which the climate system tries to remove by transporting heat through the atmosphere and the ocean.

The absorbed radiation heats the air from below, and it expands and rises. This effect is largest at low latitudes. The uprising air is replaced by air flowing at low altitudes from higher latitudes. If the Earth were not rotating, the result would be a single large convection cell with air ascending in the equatorial regions and descending in the Polar Regions. As a consequence of the Earth's rotation, however, three convection cells are formed (Hadley, Ferrel, and Polar), as shown in Fig. 13.2-1. The Earth's rotation results in the Coriolis force which causes the air masses to deviate to the right when moving in the northern hemisphere and to the left in the southern

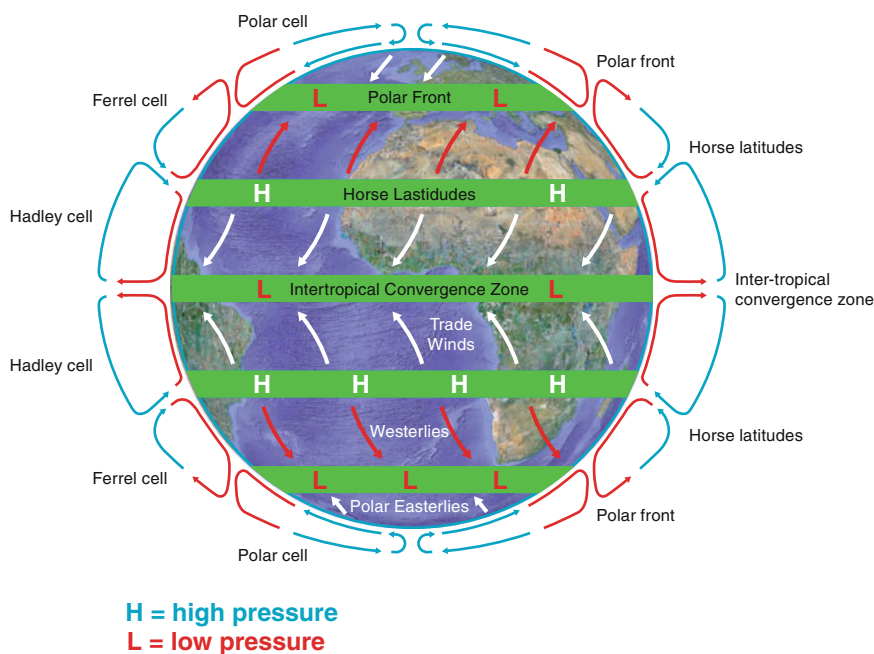


Fig. 13.2-1 Schematic representation of the tropospheric circulation. The solar energy drives three convection systems in each hemisphere. The meridionally flowing air is deviated by the Coriolis force to the *right* in the northern and to the *left* in the southern hemisphere. Latitudinal bands with uprising air are characterized by low-pressure systems and high precipitation rates, bands where the air descends by high pressure and low precipitation

hemisphere. The latitude bands where the air rises (Intertropical convergence zone (ITCZ) and Polar front) are characterized by low pressures (L in Fig. 13.2-1), the bands where the air descends by high pressures (Horse latitudes, poles).

The height to which a rising air parcel moves depends on its temperature, density, and water content. A rising air parcel cools down (at the “lapse rate”) and becomes denser. As soon as its temperature and density reaches the same values as the surrounding air, it stops rising, and above this height, vertical mixing is strongly suppressed. Whereas the temperature decreases with height in the troposphere, it increases with height in the stratosphere. The boundary between troposphere and stratosphere is the tropopause which is defined as the region where the lapse rate is lower than 2 K km^{-1} .

The larger solar input at low latitudes and the latent heat released when water vapour condenses in the rising air lead to a strong latitudinal dependence of the height of the tropopause. It reaches 16–18 km ($90\text{--}120 \text{ g cm}^{-2}$) in the tropics and 8–12 km ($220\text{--}350 \text{ g cm}^{-2}$) in Polar Regions (Fig. 13.2-3). As a result of the weather patterns and the daily and seasonal changes in insolation, the height of the tropopause is variable in space and time.

Stratosphere. In contrast to the troposphere, the temperature increases with altitude in the stratosphere primarily as a result of the absorption of ultraviolet radiation by ozone. This leads to a rather stable stratification with much less vertical transport than in the troposphere.

This difference is also reflected in the residence time (the average time an aerosol or a parcel of air resides in the stratosphere). While the residence time in the troposphere varies from a few days to a few weeks, it steadily increases from the bottom of the stratosphere from 0.5 to several years (Fig. 13.2-2).

Another difference between stratosphere and troposphere is the mechanisms that give rise to the transport of air masses. In the troposphere, thermal forcing prevails

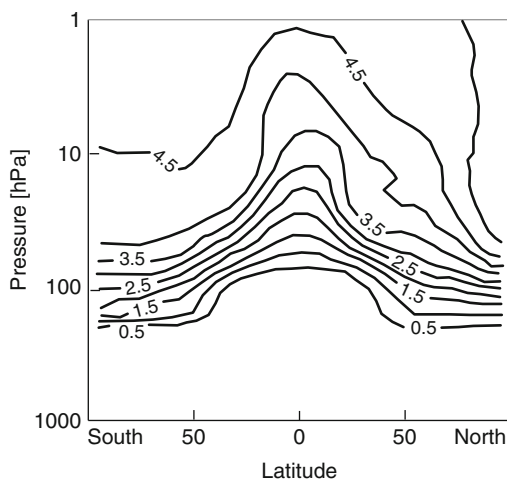


Fig. 13.2-2 Mean age of stratospheric air in years averaged zonally. Data from the Canadian Middle Atmospheric model for current conditions. For the conversion of pressure units, see Table 10.1-2. After Fig. 7 in (Shepherd 2007)

and drives the circulation in the Hadley and other cells, while in the stratosphere, waves are the dominant driving force (Fig. 13.2-3).

13.2.1 BOX Potential Temperature

The potential temperature θ of an air parcel is its actual temperature T for a reference standard pressure P_0 (usually 1,000 mb).

$$\theta = T \left(\frac{P_0}{P} \right)^{\frac{R}{C_p}} \quad (\text{B13.2.1-1})$$

θ : Potential temperature in K

T : absolute temperature in K

P : atmospheric pressure

R : gas constant

C_p : specific heat capacity for constant pressure

The concept of potential temperature is very useful for dynamic considerations. If an air parcel moves across a mountain, it cools down while rising and warms again during descent. However, the potential temperature does not change as long as the processes are adiabatic (no heat transfer, evaporation, or condensation). In other words, an air parcel can follow a line of constant potential temperature without losing or gaining any energy. A surface of constant potential temperature is sometimes called an “isentropic surface”.

Figure 13.2-3 shows a schematic of a hemisphere from the equator to the pole (Holton et al. 1995). The thick line depicts the tropopause between the troposphere and stratosphere, which ranges between ~300 mb at the pole and ~120 mb at the equator. The height of the tropopause, which is fluctuating and seasonally dependent, has important implications on how long the produced cosmogenic radionuclides reside in the atmosphere. Integration of the curves in Fig. 10.3.3-1 shows that near the equator, ~40% of the production occurs in the stratosphere. The lower tropopause in polar regions means that ~95% of the production occurs in the polar stratosphere. The differences in atmospheric circulation above and below the tropopause have an important influence upon the deposition of the cosmogenic nuclides in the archives. The thin solid lines represent isentropic or constant potential temperature surfaces in kelvins (see Box 13.2.1), which play an important role in the transport of the cosmogenic nuclides to the surface. An air parcel can move along these surfaces, neither gaining nor losing energy. The line at 380 K represents the lowest isentropic line which lies fully in the stratosphere, while the line at 300 K is the highest isentropic line fully in the troposphere. The isentropic lines between these limits cross the tropopause at mid-latitudes and provide a gate through the barrier of the tropopause. The tropics are the main gate for transport from the troposphere into the stratosphere where the responsible

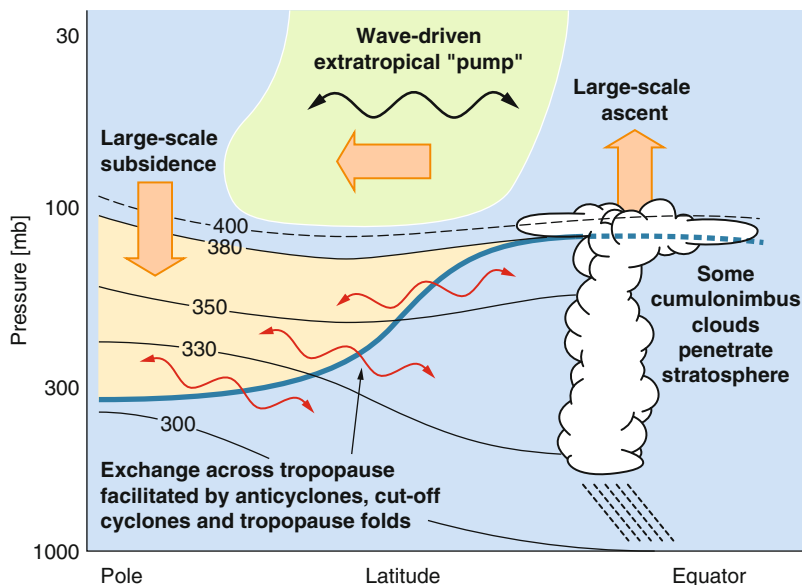


Fig. 13.2-3 Dynamics of the stratosphere and the stratosphere–troposphere exchange. The *thick line* represents the tropopause, thin lines isentropic surfaces. Wavy double-headed arrows denote meridional transport by eddy motion. The *broad arrows* depict the global transport by the wave-driven Brewer–Dobson circulation. For the conversion of pressure units, see Table 10.1-2. After (Holton et al. 1995)

process is a combination of overshooting of convective clouds and large-scale rising.

The broad arrows in the stratosphere represent the large-scale transport of air, known as the Brewer–Dobson circulation. This circulation is driven by waves propagating up from the troposphere. They transfer energy and angular momentum to the stratosphere and in combination with the Coriolis force, they result in a pumping action (“the extratropical pump”) that transfers mass flux towards the poles (Shepherd 2007), followed by subsidence to lower heights. The loop is closed by the mass flux across the tropopause at mid-latitudes along the surfaces of constant potential temperature. The Brewer–Dobson circulation occurs mainly in the winter hemisphere. During the absence of solar heating, the temperature in the polar regions drops dramatically, causing a strong temperature gradient and an eastward flow of air around the poles called the polar vortex. The Brewer–Dobson circulation is the result of a complex interplay between radiation, planetary waves, and subsidence processes in the polar vortex. It ranges up into the mesosphere (above 80 km) and transports the air masses from the tropics into the polar regions. As a consequence, the oldest air masses with ages of 6–10 years are found in the polar stratosphere.

Another important difference between the stratosphere and the troposphere is the low water vapour concentration. During the convective rise of the tropical air masses to the tropopause (up to 20 km), the temperature drops to about -80°C . As a consequence, water and many other substances are removed by rainfall in the lower troposphere, and by condensation and freezing processes at higher altitudes. As discussed below, this results in efficient scavenging of the cosmogenic radionuclides in the equatorial zone.

The greatest production rate of the cosmogenic radionuclides occurs in the vicinity of $100\text{--}200\text{ g cm}^{-2}$ in the lower stratosphere at high latitudes (Fig. 10.3.3-1). Figure 13.2-2 shows that some of them will then remain in the stratosphere for years. Figure 13.2-3 shows that the remainder of the high latitude stratospheric production can follow the isentropic lines through the “gate” into the troposphere, and then be distributed by the tropospheric circulation. Clearly, the transport processes are complex, and involve the interaction of a number of different physical processes operating on different time scales. We will return to this later.

The tropopause plays another especially important role for cosmogenic radionuclides, as a barrier between the stratosphere and the troposphere. The reason is that $\sim 65\%$ of the cosmogenic radionuclides are produced in the stratosphere, while the source of most other important atmospheric constituents such as particles and aerosols lies in the troposphere. It is also important to note that the tropopause is subject to strong spatial and temporal variability superimposed upon the decreasing trend with increasing latitude, as indicated in Fig. 13.2-4. For example, it rises during periods of warmer climate. All these lead to a major source of variability, or “noise” in the system response of the atmosphere.

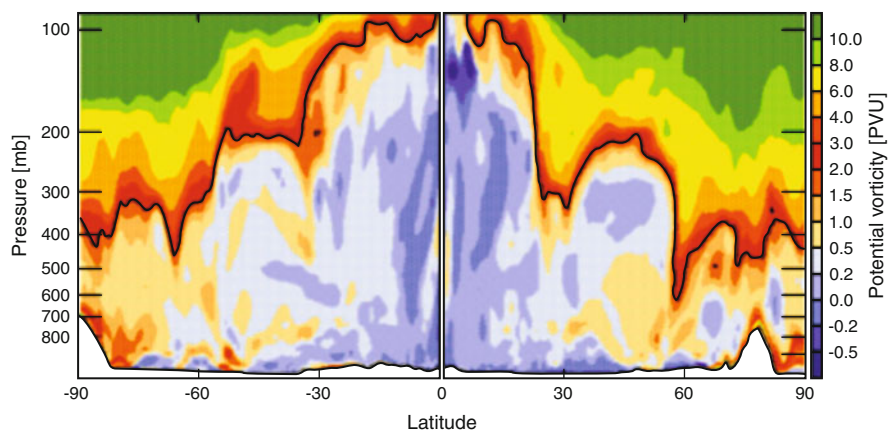


Fig. 13.2-4 Height of the tropopause (*black line*) along the longitude of 55°W . The complex structure and its association to the atmospheric circulation are obvious. Note the asymmetry between the hemispheres. (after H. Wernli, ETH Zurich)

13.3 Removal or Scavenging Processes

As discussed in Chap. 10, the cosmogenic radionuclides are formed as single atomic nuclei, and immediately combine with electrons to become neutral atoms. After that, their fate depends strongly on their chemical properties. Some will react with oxygen (Be, Al, C) or hydrogen (Cl). Many have a strong tendency to become attached to aerosols (Be, Al), while others may stay in a gaseous form (C, Cl). ^{14}C forms $^{14}\text{CO}_2$ and enters the global carbon cycle. ^{36}Cl forms H^{36}Cl or $^{36}\text{Cl}_2$.

Independent of their exact chemical form, the transport in the stratosphere is mainly determined by the flow of the air masses except when the aerosol particle grows very large and gravitational settling begins. On the contrary, the chemical form of the cosmogenic radionuclides plays a much more important role in the troposphere where scavenging takes place. The incorporation of cosmogenic radionuclides into water droplets is a complex dynamic process that depends on many variable parameters, as briefly outlined below.

In the previous section, we discussed how air masses and aerosols with their attached cosmogenic radionuclides are transported within the atmosphere. To be stored in an archive, the cosmogenic radionuclides must then be removed from the atmosphere. These processes occur preferentially in the lower part of the troposphere where clouds are formed. We can distinguish three different processes that are responsible for what is often referred to as “scavenging” in the literature describing the removal of aerosols from the atmosphere.

13.3.1 Wet Deposition

“Wet deposition” includes all processes which involve water in the form of rain droplets, snowflakes, or hail, and is the most efficient removal process. In the lower troposphere, where the water content is high and clouds are formed, the probability that a small aerosol particle (carrying Be or Al) or a gaseous molecule (HCl , CO_2) will collide with a water droplet is high. Each interaction increases the probability that the particle or the molecule will become incorporated into the droplet. A raindrop will typically carry several thousand aerosol particles to the ground. The efficiency of this “in-cloud scavenging” process is well known from the fact that atmospheric visibility is often much better after rain. Some aerosols can also be picked up by the falling raindrops below the clouds; however, this “below-cloud scavenging” process is much less important.

The wet deposition removal process can be described by the differential equation:

$$dN_s/dt = -\lambda_{sp}N_s, \quad (13.3-1)$$

where

N_s : number of aerosol particles scavenged per unit volume and size range s ,

λ_{sp} : washout coefficient reflecting the probability that an aerosol particle in the size range s gets removed for a precipitation rate p .

Figure 13.3-1 [Fig. 5.8 from (Greadel and Crutzen 1993)] shows the experimentally determined dependence of the washout coefficient on the diameter of the aerosol particles. Due to their comparatively much larger surface, snowflakes have a factor of 10–100 higher washout coefficient. Theoretical considerations lead to the dependence indicated by the dashed line with a distinct minimum at about 1 μm . Experimental data confirm the theoretical expectations. As a general rule, washout or wet deposition is the main removal process for aerosols from the atmosphere if the mean annual precipitation rate is larger than about 10 cm.

It is clear from the above discussion, and Eq. (13.3-1), that wet scavenging is most efficient in cloudy regions, and where rainfall is high. This is further demonstrated by Fig. 13.3-2, which shows clearly that wet precipitation tends to be higher over the equatorial and mid-latitude oceans. Thus, the cosmogenic radionuclides produced in the equatorial troposphere, together with those that have arrived there through the “gate” from the stratosphere, will have a high probability of being rapidly removed from the atmosphere at low latitudes. Removal will also be rapid in the vicinity of the “polar front”, partly due to the rising air (and cloudiness) of the Polar and Ferrel cells (Fig. 13.2-1) and partly due to the factor of 10–100 higher scavenging efficiency of snowflakes. On the contrary, the low precipitation rates of the Horse Latitudes ($\sim 30^\circ\text{N}$ and S) implies low scavenging rates of both the locally produced nuclides and those that have reached there through the gate in the tropopause.

Consider now the cosmogenic radionuclides that remain in gaseous form. HCl is highly soluble in water, and consequently, it is quickly scavenged when a molecule

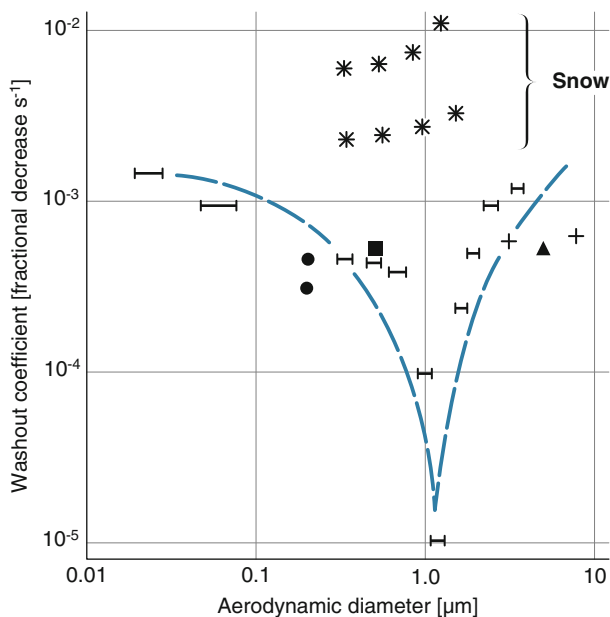
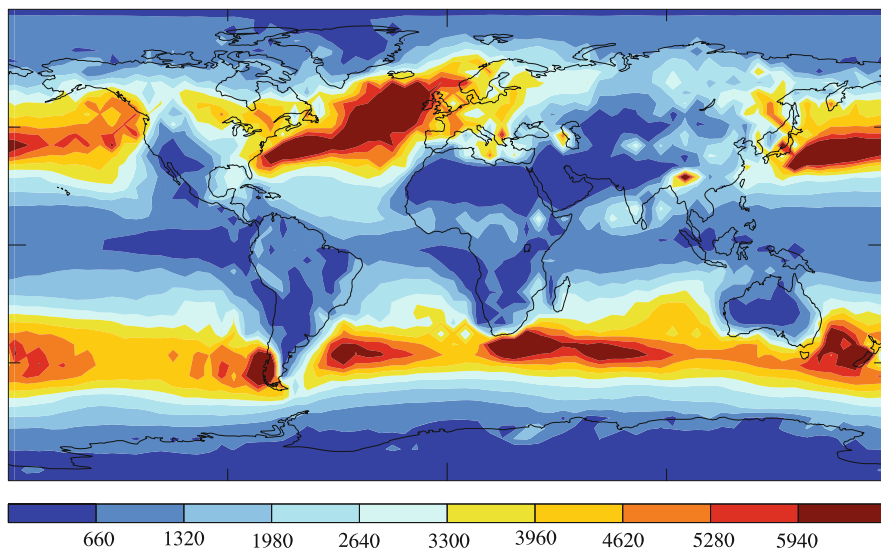
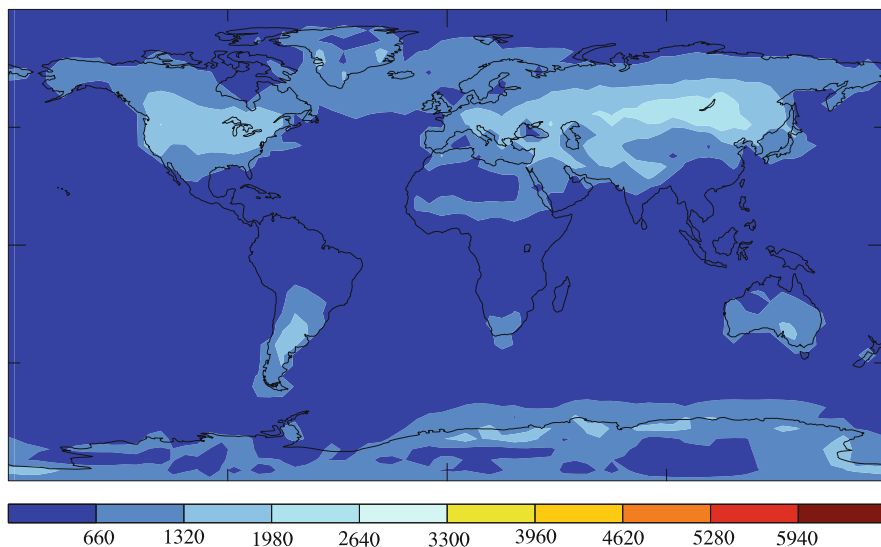


Fig. 13.3-1 Dependence of measured washout coefficients for snow (stars) and rain (other symbols) on the particle size. Dashed lines show the trend in the rain data. (After (Greadel and Crutzen 1993) Fig. 5.8)

a Annual mean wet ^{10}Be deposition ($10^{-27} \text{ kg/m}^2/\text{s}$)**b** Annual mean dry ^{10}Be deposition ($10^{-27} \text{ kg/m}^2/\text{s}$)**Fig. 13.3-2** Computer model showing the worldwide distribution of “wet” and “dry” mean annual deposition of ^{10}Be . After (Field et al. 2006)

collides with a water droplet. In general, its transport in the troposphere is similar to that of the cosmogenic radionuclides attached to aerosols. The case of $^{14}\text{CO}_2$ is quite different. It is much less soluble in water and tends not to be scavenged by

clouds or rain. It remains in the atmosphere in a gaseous form for ~ 7 years, during which time it becomes well mixed throughout the global atmosphere (see Sect. 13.5.3). As a consequence, the ^{14}C in the biological archives represents a global average of the production rate, while the ^{36}Cl archive in polar ice represents a more restricted range of latitudes, similar to those of ^{10}Be and ^{26}Al .

13.3.2 Dry Deposition

Dry deposition means that an airborne particle adheres when it touches any surface such as the pedosphere (soil) or vegetation. The deposition flux for a specific species of particles i can be described as

$$F_i = C_i v, \quad (13.3-2)$$

where:

C_i : concentration of particles of species i at a reference height above ground,
 v : deposition (vertical) velocity.

v (and therefore F_i) depends on the particle size and ranges from 0.02 to 2 cm s^{-1} . The chemical properties determine how easily the particle sticks to the surface, while the size is important for the dynamics. Very small particles move faster and have larger diffusion constants, while large particles have a large momentum which prevents them from following the rapid deflections of the air mass near the surface, thereby reducing the probability of forceful impacts. Experimental data show that the particles with diameters between 0.1 and 1 μm exhibit the lowest deposition velocity of about 0.02 cm s^{-1} . For diameters of 10 μm , it rises to 1 cm s^{-1} .

Dry deposition is the dominant sink process in areas with very low precipitation rates ($< 10 \text{ cm year}^{-1}$) such as in deserts and the inner part of Antarctica (Fig. 13.3-2).

13.3.3 Gravitational Settling

If a particle grows large, then gravitational settling has to be considered. According to Stoke's law, the settling velocity v_s of a small spherical particle with radius r and density ρ_p in a fluid like air with dynamical viscosity μ , density ρ_f , and gravitational acceleration g is

$$v_s = \frac{2}{9} \frac{r^2 (\rho_p - \rho_f) g}{\mu} C_c \quad (13.3-3)$$

C_c is a correction factor that is proportional to the mean free path of air. This and the relatively long residence time mean that gravitational settling mainly plays a role in the stratosphere.

13.3.4 *The Big Picture*

In summary, as a result of the relatively large mean global precipitation rate of about 1 m year^{-1} , wet deposition is by far the main removal mechanism for aerosols carrying cosmogenic radionuclides from the atmosphere. Dry deposition only plays a significant role in very dry regions such as Antarctica and the deserts. Gravitational settling occurs mainly in the stratosphere where the vertical transport is slow and the residence times are long.

Ice cores from the polar regions are important archives for many cosmogenic radionuclides, and the above discussion illustrates the complex question of how the different regions of the earth and atmosphere contribute to those archives. We can recognize the following components:

- (a) Production in the polar troposphere that is scavenged into the polar region within several days to several weeks.
- (b) Stratospheric production at all latitudes that has been mixed by the Brewer–Dobson circulation, and after a residence time of >0.5 years, has entered the troposphere at mid-latitudes and has then escaped being scavenged en route to the high latitude archive.
- (c) Production in the equatorial and mid-latitudinal troposphere that has escaped scavenging en route to high latitudes over a period of weeks.
- (d) Gravitational settling from the polar stratosphere, through the tropopause, leading to relatively rapid sequestration of the nuclides in the polar archive.

From the above, and from Fig. 13.3-2, it is clear that the transport and scavenging processes will depend to some extent upon the distribution of oceans and land between the equator and high latitudes. For this reason, it is not a priori necessary that the atmospheric averaging processes will be the same for Greenland and Antarctica.

Note also the difference between ^7Be and ^{10}Be . The 53.2-day half-life of ^7Be is short compared to the residence time in the stratosphere (Fig. 13.2-2), and roughly comparable to the residence time in the troposphere. Consequently, component (b) above will be devoid of ^7Be , while a greater fraction of the ^7Be in component (c) will probably have decayed than in the case of component (a). These differences can be used to investigate atmospheric processes as we discuss in Chap. 19.

It is therefore clear that the properties of the averaging processes, and of the system noise, are quite complex and dependent upon location and time, and that there will be no simple way to estimate the relative importance of the four

components (a) to (d) above. This is only possible through the use of modern mathematical models of the atmospheric circulation processes, as we discuss in the next section.

13.4 Modelling the Atmospheric Transport

As discussed above, the atmosphere is a very dynamic and complex system which is driven not only by solar radiation but also by Coriolis forces due to the Earth's rotation, and by various wave and chemical processes. It is therefore very difficult to predict how much the production signal of cosmogenic radionuclides measured, for example, in an ice core in Greenland has been influenced by atmospheric mixing and scavenging, or by long-term climate change. The only way to answer this question is to use mathematical models that approximate reality.

Historically, the first very simple models were the so-called box models. They divide a system such as the atmosphere into a few boxes, and assume that each box is homogeneously mixed and that it exchanges with the surrounding boxes with a flux that can be expressed as a mathematical equation. For example, the atmosphere could be described by a 2-box model consisting of the stratosphere and the troposphere. Although this is an extreme simplification, it provides a useful global picture and has been successful in describing the ^{36}Cl fallout of the nuclear bomb tests in the 1950s and 1960s of the twentieth century (Synal et al. 1990). In Sects. 14.7.1 and 14.7.2, we use two box models to discuss the ^{10}Be and ^7Be in the atmosphere and ^{10}Be and ^{26}Al in deep-sea sediments, respectively. To make the model more realistic, the number of boxes can be increased (Scheffel et al. 1999). For example, this is important to interpret correctly the observed ^{10}Be variations in the polar regions where the geomagnetic modulation is zero and the solar modulation largest.

A fully realistic model needs to include the physics (energy, momentum, etc.), the different circulation regimes in Fig. 13.2-3, the topography of the Earth, the geographical distribution of scavenging processes, and the coupling with the oceans. This type of model is called a general circulation model (GCM). The spatial resolution in present-day GCMs is too coarse to model the stratosphere–troposphere exchange properly, and many processes are not yet fully implemented. In addition, there are specific atmospheric models which are designed to describe the atmospheric processes in more detail; however, since they are not coupled to the global system, they can only be used on short time scales. In general, we can anticipate that GCMs will greatly improve our understanding of the transportation and sequestration processes in the near future.

At the time of writing, there have been two independent sets of GCM calculations of the world-wide precipitation of ^{10}Be , and comparison of the results with measured data looks very promising as far as the cosmogenic radionuclides are concerned (Field et al. 2006; Heikkilä et al. 2008a, b). We briefly summarize these GCM results here, and use them to illustrate the qualitative concepts discussed in Sects. 13.2 and 13.3.

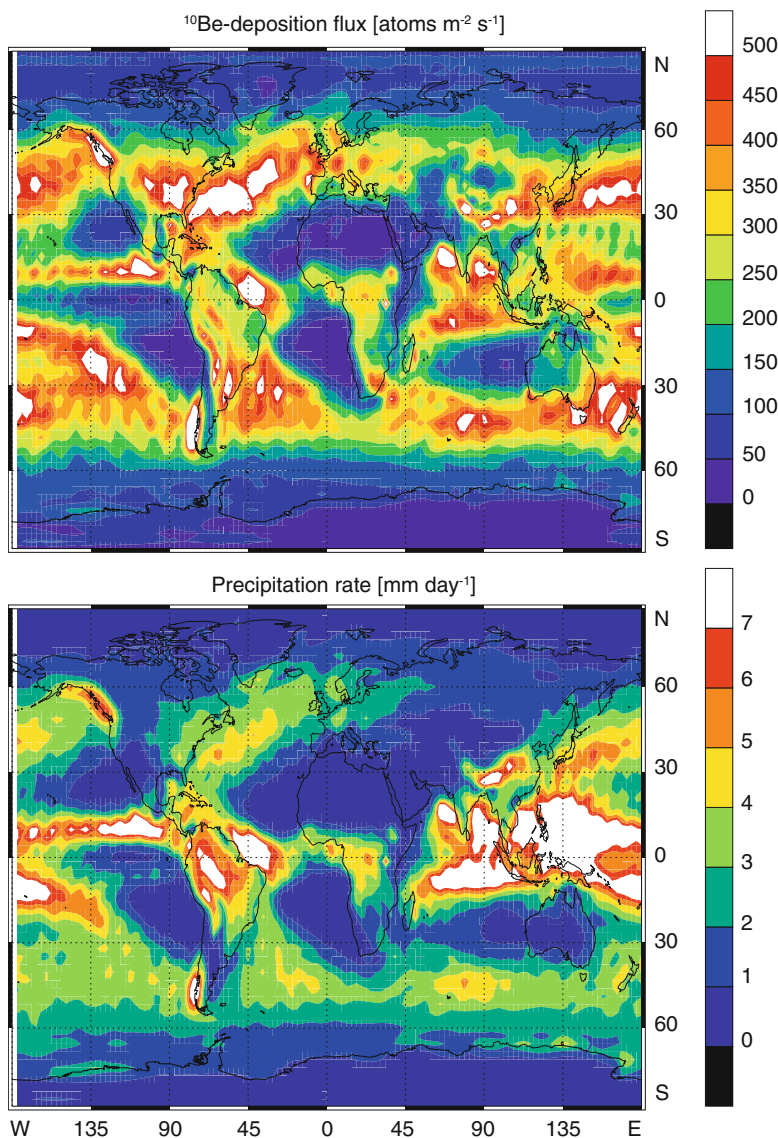
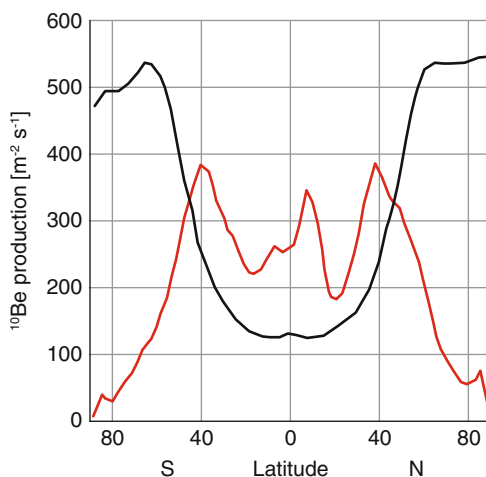


Fig. 13.4-1 Total ^{10}Be deposition flux in atoms $\text{m}^{-2} \text{s}^{-1}$ (upper panel) and total precipitation rate in mm per day (bottom panel). After (Heikkilä 2007)

Figure 13.4-1 shows the total ^{10}Be deposition flux (upper panel) together with the total precipitation rate (lower panel). This figure demonstrates clearly that the flux varies by more than a factor of ten, worldwide, and that it is least in the polar regions, where the cosmogenic production rate is greatest (Fig. 10.3.3-2). Closer inspection shows that the ^{10}Be deposition flux is governed by two main factors: the

Fig. 13.4-2 Comparison of the zonally averaged latitudinal production rate of ^{10}Be (black line) with the modelled deposition rate (red line) for the period 1986–1990. Note that the production pattern caused by the geomagnetic field is completely changed by the transport and deposition processes. After (Heikkilä et al. 2009)



precipitation rate which peaks in the tropics and at mid-latitudes, and the stratosphere–troposphere exchange which also peaks at mid-latitude. The original latitudinal distribution with low ^{10}Be production in the tropics and high production in the polar regions is completely changed by the transport effects. This is vividly demonstrated when both production and deposition rates are zonally averaged (Fig. 13.4-2).

The production rate (black line) shows the well-known latitudinal dependence, which is due to the screening properties of the geomagnetic field (see Fig. 10.3.3-2). The slight decrease at high latitudes in the southern hemisphere is due to the high altitude of Antarctica (4,000 m) which reduces the production rate integrated through the atmosphere. The deposition rate (red line) shows a completely different pattern, with peaks in the mid-latitudes and in the tropics. The deposition fluxes in the Polar Regions are comparatively low. This pattern partly reflects the stratosphere to troposphere exchange that takes place in the mid-latitudes (through the “gate” in the tropopause) and the fact that about 65% of the cosmogenic radionuclides are produced in the stratosphere. It is also influenced by the geographical distribution of the scavenging processes; thus, the equatorial and mid-latitude peaks reflect the higher precipitation at those latitudes, while the troughs near $\sim 20^\circ$ reflect the lower precipitation in the “Horse Latitudes”.

Figure 13.4-3 shows the global distribution of the stratospheric fraction of the produced ^{10}Be . In view of the long residence time in the stratosphere (Fig. 13.2-2), it can be assumed as a first approximation that this ^{10}Be is well mixed between all latitudes. It is interesting to note that everywhere on the globe the stratospheric fraction of the total ^{10}Be deposition flux is more than 50%, the lowest values being found in the tropics and the highest in the mid-latitude regions where the exchange between stratosphere and troposphere takes place. In Polar Regions, the contribution is between 65 and 70% (Greenland) and between 60 and 65% (Antarctica) respectively. The largest deposition rates are expected at mid-latitudes high up in

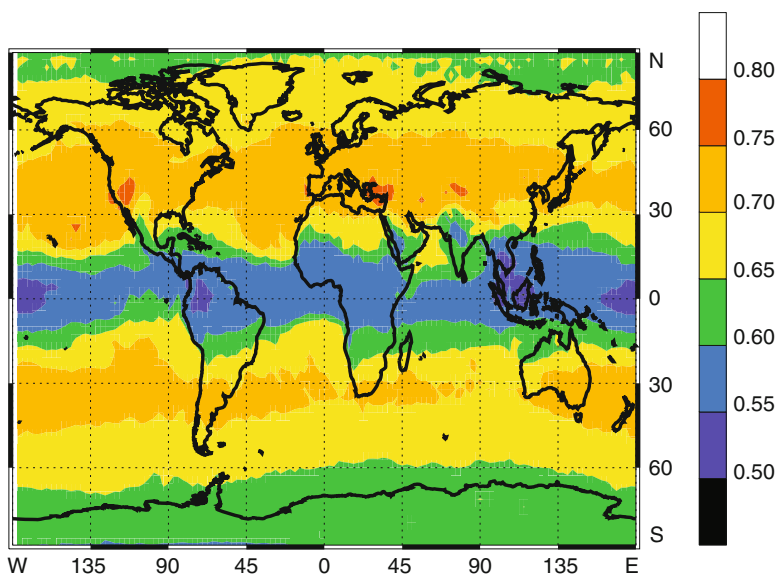


Fig. 13.4-3 Modelled contribution of the fraction of ^{10}Be produced in the stratosphere to the total ^{10}Be deposition (Heikkilä et al. 2009). The largest fraction is found in the mid-latitudes as a consequence of the stratosphere to troposphere exchange at those latitudes. The contributions in the Polar Regions range between 60 and 70% with higher values in Greenland than in Antarctica

the mountains (Rocky Mountains, Himalayas). Although the latitude gradients in precipitation are the largest, there is also a significant longitudinal variability.

Figure 13.4-4 examines the degree of transport, and mixing that occurs before the ^{10}Be is scavenged to Earth. In Fig. 13.4-4a, the ultimate destination of the (well-mixed) ^{10}Be produced in the stratosphere is shown as a function of latitude after integrating over all longitudes. There is a small asymmetry between the two hemispheres, which is reflected in a slightly larger peak in northern mid-latitudes, and a small peak in the tropics. This asymmetry can probably be attributed to stronger planetary waves above the continents (primarily in the northern hemisphere), enhancing the stratosphere to troposphere exchange and increased scavenging due to a larger tropical precipitation rate in the northern hemisphere.

Figure 13.4-4b shows the mean latitudinal deposition of the ^{10}Be produced in the troposphere between 0° and 30° in both hemispheres. The total flux is considerably smaller than the stratospheric flux in (a) because the atmospheric production volume is much smaller, and the production rate at low latitudes is reduced due to the geomagnetic shielding (Fig. 13.4-2). Very little of the ^{10}Be produced in the tropical troposphere is transported to the polar latitudes. The asymmetric shapes of the precipitation curves indicate that the equator acts as a strong barrier between the hemispheres. The small peak in the northern tropics (Fig. 13.4-2) is probably due to the higher tropical precipitation rate of the northern hemisphere.

In Fig. 13.4-4c, the latitudinal deposition is shown for ^{10}Be stemming from production in the troposphere between 30° and 60° latitude. As a result of the transport

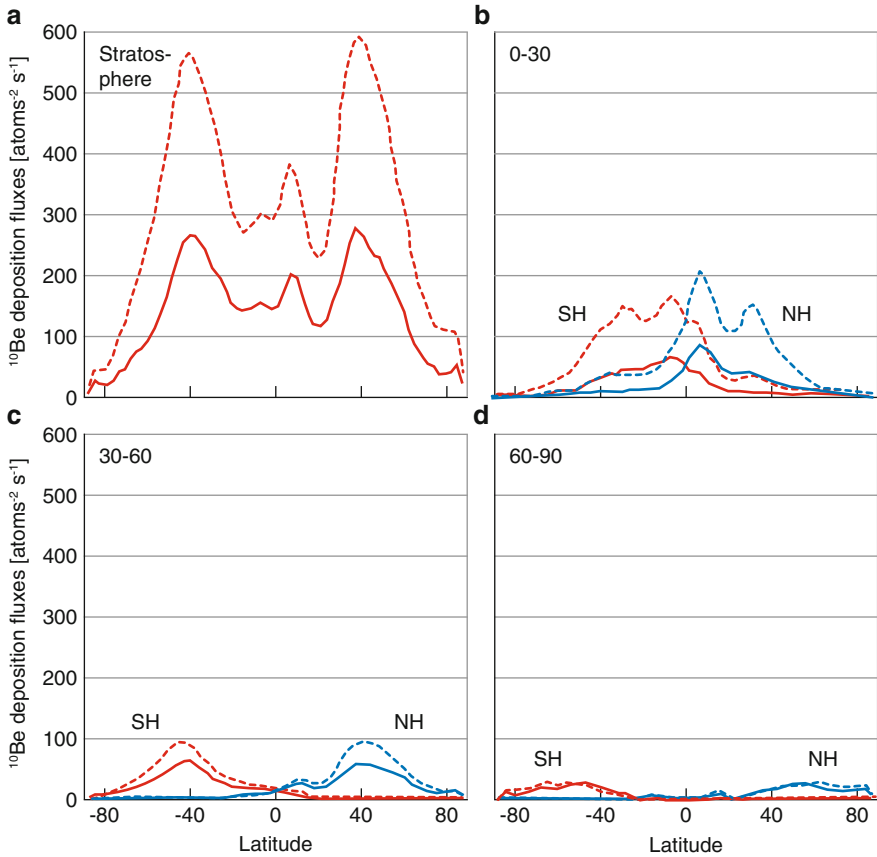


Fig. 13.4-4 Zonal means of the ^{10}Be deposition fluxes that originated in seven atmospheric source regions, for the period 1986–1990 (*solid line*), and the Laschamp event (*dashed line*) when the geomagnetic field was close to zero. (a) The deposition flux of the stratospheric fraction; (b) the deposition fluxes of the ^{10}Be produced in the tropical troposphere ($0\text{--}30^\circ$) for both Northern (NH) and Southern (SH) Hemisphere; (c) tropospheric mid-latitude fraction ($30\text{--}60^\circ$); and (d) tropospheric polar fraction ($60\text{--}90^\circ$). Units are $\text{atoms m}^{-2} \text{s}^{-1}$.

processes, the ^{10}Be is spread over all latitudes between the poles and the equator; however, the main deposition (58–60%) remains in the mid-latitudes where it was produced. Virtually none of the ^{10}Be produced in one hemisphere reaches the other.

Finally, Fig. 13.4-4d shows that there is substantial transport of the ^{10}Be from high latitudes ($60^\circ\text{--}90^\circ$) to the mid-latitudes, resulting in significant flattening and dampening of the peak.

The results in Fig. 13.4-4 for the present epoch are presented in relative form in Fig. 13.4-5. Each column represents the production in one of the atmospheric compartments used in Fig. 13.4-4, expressed as 100%. The different areas within the columns show the percentage of the ^{10}Be produced in this compartment that is transported to the specified latitudinal zone on Earth. For example, Fig. 13.4-5 tells

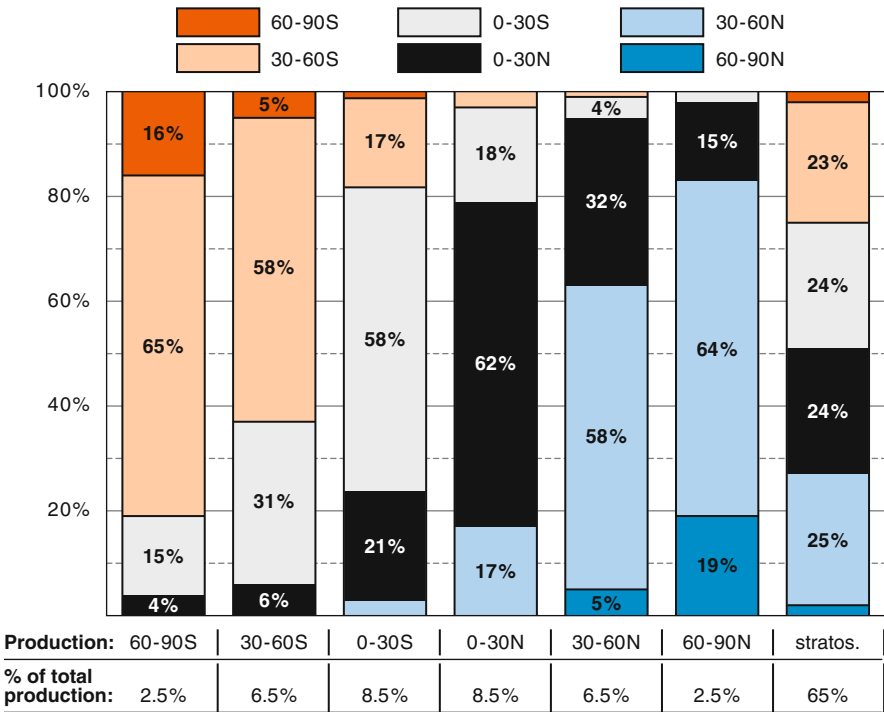


Fig. 13.4-5 Relative distribution of ^{10}Be produced in each of seven atmospheric compartments (Troposphere: $60^{\circ}\text{--}90^{\circ}\text{S}$, $30^{\circ}\text{--}60^{\circ}\text{S}$, $0^{\circ}\text{--}30^{\circ}\text{S}$, $0^{\circ}\text{--}30^{\circ}\text{N}$, $30^{\circ}\text{--}60^{\circ}\text{N}$, $60^{\circ}\text{--}90^{\circ}\text{N}$, and Stratosphere) into the specified tropospheric compartments (see legend at top)

us that 58% of the ^{10}Be produced in the tropospheric compartment $30^{\circ}\text{--}60^{\circ}$ south is deposited in the same compartment, and only 5% is transported to the zone $60^{\circ}\text{--}90^{\circ}$ south. At the bottom, the percentage of the total worldwide production is given for each compartment (Heikkila et al. 2009).

It is interesting to note that roughly 60% of the ^{10}Be is deposited in the tropospheric compartment in which it has been produced and that the stratospheric production is equally distributed to the tropospheric compartments (~24% each) except for the polar regions (4% in total).

Having determined the global destinations of the ^{10}Be produced in the seven production regions, we now turn the problem around and determine the relative contributions to a specified archive; we consider the Antarctic archive as an example. From the right hand column of Fig. 13.4-5, we see that 2% of the 65% stratospheric production is deposited in the southern polar cap, corresponding to 13 permil of the total global production (1 permil, symbol ‰, = 0.1%). Likewise, the tropospheric contributions from between $30^{\circ}\text{--}60^{\circ}\text{S}$ and $60^{\circ}\text{--}90^{\circ}\text{S}$ are 3.3 and 4.0 permil respectively. That is, when we sample ^{10}Be in Antarctic ice, the greater proportion (64%) has come from the stratosphere, 16% from the troposphere at mid-southern latitudes, and 20% from the polar region troposphere. The

contributions from the equatorial region, or from the northern hemisphere are negligible. For the arctic archive, the numbers are 60%, 18%, and 22%.

In summary, approximately 60–64% of the ^{10}Be in the polar archives have come from the well-mixed stratosphere, representing the global average of the production rate. The remaining 36–40% has originated in the troposphere, approximately half being due to production in the polar cap itself, and half at intermediate latitudes. In total, the amplitude of the heliospheric variations in the polar archive is close to the global average. This result is supported by the comparison of ^{10}Be from Greenland with the production rate derived from the $\Delta^{14}\text{C}$ tree ring record in Fig. 17.3.2-1.

Referring again to Fig. 13.4-4, the dotted lines are the results of GCM calculations in which the geomagnetic field is assumed to be zero, with the result that “polar production” extends to the equator. That means, while the production at high latitudes is unchanged, near the equator it has increased by a factor of four (Fig. 10.3.3-4). Nevertheless the general features of the curves in Fig. 13.4-4 are unchanged, the peaks and troughs occurring at the same latitudes as for the normal geomagnetic field (solid line). That is, the precipitation is controlled by the geometric features of the atmospheric transport and scavenging processes which were assumed to be the same as today in this experiment.

13.4.1 Summary

As a result of the nuclear reactions between cosmic rays and the atmosphere and the geomagnetic shielding effects, the production rate of cosmogenic radionuclides in the atmosphere is strongly altitude and latitude dependent. In addition, changes in solar activity cause much larger production variations in the polar regions than near the equator, while changes in the geomagnetic field result in large production changes at low latitudes, and no effect whatsoever in the polar regions.

From this it would appear that an archive in the tropics would be ideal to study the history of the geomagnetic field, and one near the poles to study solar activity in the past. However, examination of the transport processes has provided a completely different picture. The atmosphere is a highly dynamic system consisting of a thermally driven troposphere with three large convection systems in both hemispheres with residence times ranging from a few days to several weeks, and a rather stratified stratosphere with residence times of a few years. The wave-driven Brewer–Dobson circulation causes some mixing within the stratosphere and is mainly responsible for air exchange between stratosphere and troposphere in the mid-latitudes. As a consequence of these dynamical processes, there is substantial mixing of the atmosphere with the result that the flux of ^{10}Be into a polar archive is rather well mixed, and the production signal imprinted on ^{10}Be by the Sun and the geomagnetic field is close to the global average.

For temporal resolutions of 1 year or even one month, the variations in the sequestration rate becomes increasingly dominated by transport processes. One has

to keep in mind that as a result of the different atmospheric residence times, cosmogenic radionuclides produced at the same time in the atmosphere have different travel times ranging from weeks to a few years before they are stored in an archive. In addition, for long time scales of millennia and more, climate change may affect the atmospheric circulation and the precipitation rate. As a result, the local deposition fluxes may change without any changes in the production rate. A more detailed discussion of the effects of a changing climate on concentration and fluxes of cosmogenic radionuclides is given in Sect. 19.4.

13.5 Geochemical Cycles

13.5.1 Introduction

In the context of the cosmogenic radionuclides, we consider a geochemical cycle to be the path followed by an element in the environmental system. After production, a cosmogenic radionuclide is subject to different transport processes depending on where it is produced and its geochemical properties. An important aspect, namely the transport from the atmosphere where most of the cosmogenic radionuclides are produced, into an archive where they are stored, has already been discussed in Sect. 13.4. In that case, the main emphasis is on estimating the production rate and its changes in the past, and the transport is often considered as a source of noise which needs to be suppressed as much as possible. However, cosmogenic radionuclides can also serve as excellent tracers to study environmental processes and in such cases, the former noise becomes the signal, and the production rate is assumed to be known as a function of time.

As we will see, the geochemical properties of a cosmogenic radionuclide can be very different depending on its chemical reactivity (inert to very reactive) and its phase (solid, liquid, or gaseous). The phase depends on the environmental conditions and also on the compounds that the nuclide forms. For example, chlorine can become attached to an aerosol in the form of HCl, or remain in gaseous form as Cl₂. Iodine can have different oxidation states which behave differently in nature. Most geochemical cycles are very complex and not yet fully understood. While this poses a problem for the interpretation of cosmogenic radionuclide records, at the same time it opens a large field of applications in which they are used as tracers.

In the following, we discuss the geochemical cycles of Be, C, Cl, and I.

13.5.2 The Beryllium Cycle

Beryllium is a light metal with some special physical characteristics: its melting point (1,560 K), specific heat (1,925 J kg⁻¹ K⁻¹), and thermal conductivity (216 W m⁻¹ K⁻¹)

are very high. It has a low density (1.85 g cm^{-3}) and a very low coefficient of linear thermal expansion ($11.4 \times 10^{-6} \text{ K}^{-1}$). Its scattering cross-section is large for high-energy neutrons and very small for thermal neutrons.

The only stable isotope of Beryllium is ^9Be . In nature it is found only in minerals combined with other elements. Beryllium is relatively rare in the universe because it is not formed in stellar nucleosynthesis (^8Be is formed in nucleosynthesis; however, it is unstable and decays into two ^4He nuclei). The ^9Be in the solar system and on Earth was formed by cosmic ray interactions with interstellar dust and has an abundance of 4–6 ppm in the Earth's lithosphere.

Chemically, Beryllium exhibits the +2 oxidation state and has a high affinity for oxygen. It attaches to particles, but similarly to Aluminium, it dissolves in warm alkali and forms $\text{Be}(\text{OH})_4^{2-}$ and H_2 gas.

The global ^{10}Be cycle is shown in Fig. 13.5.2-1. It is important to note that ^9Be and ^{10}Be have different geochemical cycles because their sources are completely different. In many cases (e.g. in ice cores), the ^9Be content is much too low to allow chemical procedures to be used to extract the Beryllium, and since we measure the

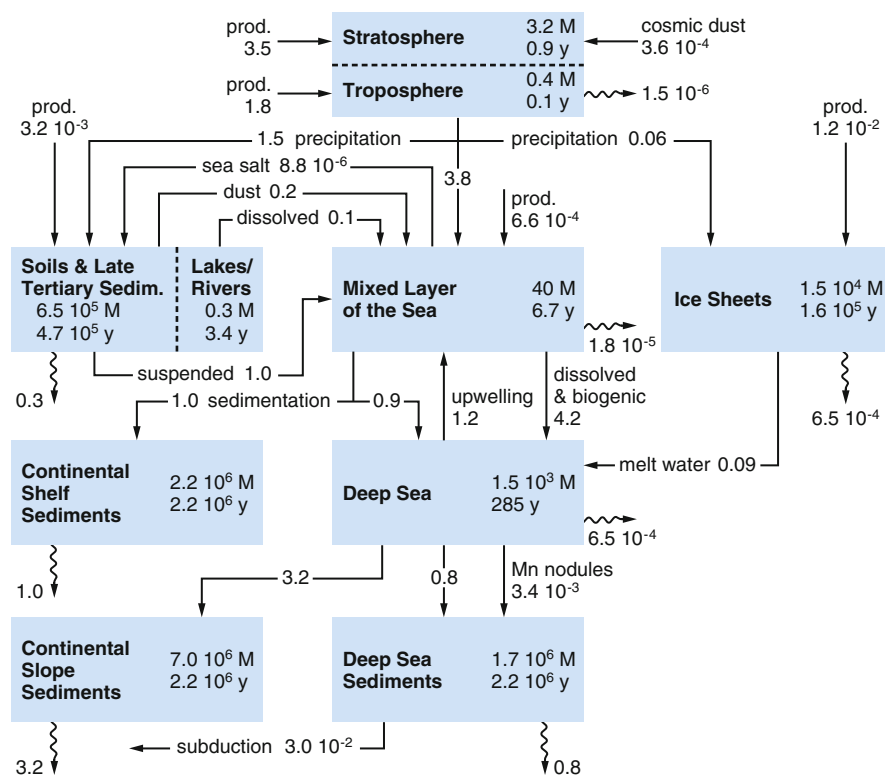


Fig. 13.5.2-1 The global ^{10}Be cycle after (McHargue and Damon 1991). The inventories of the different compartments are given in mols, and the fluxes between them, the production rates, and the decay in mols per year. The wavy lines indicate radioactive decay

ratio of ^9Be to ^{10}Be (see Chap. 15), a known amount of ^9Be has therefore to be added to the sample before extracting the beryllium. In the case of deep-sea sediments, it is possible to extract the total Beryllium. In this case, the concentrations of both ^{10}Be and ^9Be have to be measured.

As already discussed in Chap. 10, the main production of cosmogenic radionuclides takes place in the atmosphere. The in-situ production in the continental soils, the ice sheets, and the ocean contributes less than a few percent. As expected, about 30% are deposited on the continents, while 70% enter the oceans. From the 1.5 mols of ^{10}Be deposited on the continents, 1.3 mols are carried into the sea (1.1 M by water, 0.2 M by wind). This means that most of ^{10}Be ends up in ocean sediments where it ultimately decays.

13.5.3 Carbon Cycle

The carbon cycle plays an essential role on planet Earth for several reasons. Firstly, we as well as all other living organisms on Earth are based on carbon, which accounts for about 50% of the dry weight of a human. Photosynthesis, the conversion of CO_2 into organic compounds by solar light, is the main energy source of life. Secondly, it is the most important greenhouse gas beside water vapour. Its atmospheric concentration is strongly influenced by human activities such as the burning of fossil fuel, and changing land use, and both have impacts upon the global climate.

Carbon has two stable isotopes, ^{12}C (98.9%) and ^{13}C (1.1%). In the lithosphere, the weight fraction of carbon is relatively low: 0.032%. As discussed in Chap. 10, the cosmic radiation produces a radioactive isotope, ^{14}C , primarily in the atmosphere, and it participates in all biological processes along with the stable isotopes. In so doing, it provides an ideal “clock” that allows us to determine the age of biological matter produced over the past 50,000 years.

The carbon cycle can be described in a simplified manner as follows (Fig. 13.5.3-1). Atmospheric CO_2 is partly taken up by living organisms such as trees and grasses, and consumed by herbivores. After their death, the organic compounds are oxidized and some of the CO_2 returns into the atmosphere, and some is sequestered in the soil. Another part of the atmospheric CO_2 is dissolved in the ocean where a small fraction forms CaCO_3 . This sinks to the sea floor and is incorporated in sediments which after millions of years may be subducted below the continental plates. Some of that carbonate is oxidized by heating and returned into the atmosphere in volcanic eruptions.

The ^{14}C produced in the atmosphere, as described in Chap. 10, forms $^{14}\text{CO}_2$ and enters the carbon cycle in which the residence times are large. It is, therefore, crucial to understand the basics of the carbon cycle in order to use ^{14}C as a dating tool or an environmental tracer.

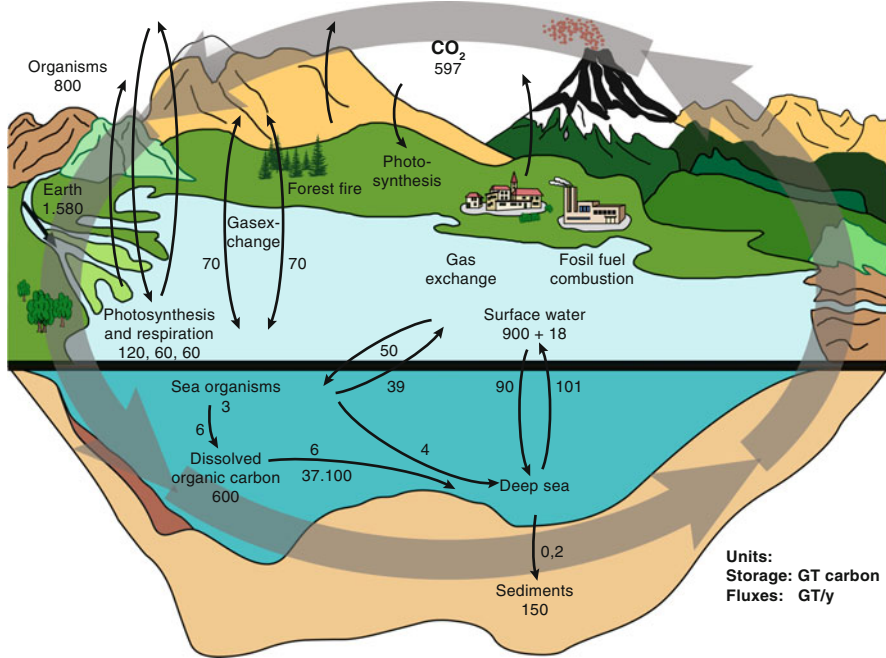


Fig. 13.5.3-1 Global carbon cycle. The carbon contents in the different reservoirs are given in Gigatons (Gt) of carbon, the fluxes (arrows) between the reservoirs in Gt per year. [After (Sturm and Lotter 1995)]

13.5.3.1 Box-Diffusion Carbon Cycle Model

It is an observational fact that ^{14}C is globally quite well mixed. Therefore, it is possible to model the ^{14}C cycle successfully with a simple box-diffusion carbon cycle model (Oeschger et al. 1975). This model is a simplification of Fig. 13.5.3-1 and consists of only 4 reservoirs (see Fig. 13.5.3.1-1). Comparison with the state-of-the-art carbon cycle models (Joos et al. 2004) including features such as a dynamic global vegetation and CaCO_3 sedimentation agrees very well with the simple box-diffusion model as long as one is only interested in the mean ^{14}C concentration of the main reservoirs.

The deep sea is treated as a diffusive medium with an eddy diffusion constant of $4,000 \text{ m}^2 \text{ year}^{-1}$. Sedimentation is neglected because each year, it only removes about 5×10^{-6} of the total carbon present in the ocean.

In the figure, N is the reservoir size in Gt (10^{15} g) of carbon, and R is the $^{14}\text{C}/^{12}\text{C}$ ratio that (for simplicity) is set artificially to 1 (100%) for the atmosphere. The years attached to the arrows give the mean residence times for this specific exchange. They are inversely related to the corresponding exchange coefficients (used later). For example, the mean residence time is estimated to be 7.7 years in the atmosphere, and therefore, the exchange coefficient between atmosphere and the mixed layer is $k_{\text{am}} = 1/(7.7 \text{ year})$. Biomass is directly formed from atmospheric CO_2 and

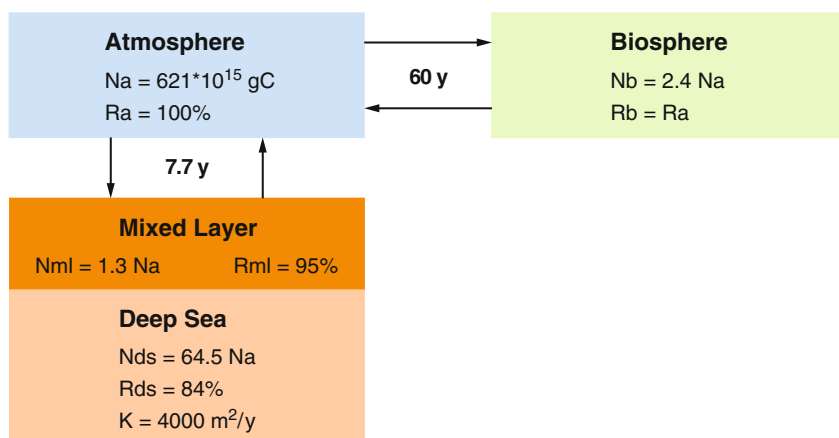


Fig. 13.5.3.1-1 Box Diffusion model after (Oeschger et al. 1975). The carbon cycle is reduced to the atmosphere (a), biosphere (b), mixed ocean layer (ml) which corresponds to the top ~70 m where the water is well mixed, and the deep sea (ds) where mixing due to the thermohaline circulation is slow. This turbulent mixing is described in the model by the eddy diffusion constant K

decomposes on average in about 60 years, which is short compared to the mean lifetime of a ^{14}C atom of 8,270 years ($T_{1/2} = 5730$ years). As a consequence, the $^{14}\text{C}/^{12}\text{C}$ ratio is approximately the same in the biosphere and the atmosphere ($R_b = R_a$), other than a small change as a result of fractionation due to the slightly greater mass of the ^{14}C atom. In practice, a correction for the fractionation can be made by measuring the $^{13}\text{C}/^{12}\text{C}$ ratio (Stuiver and Polach 1977) See Box 23.2.1.1.

Observations indicate that R for the mixed layer is 95% of that in the atmosphere. This decrease of 5% indicates a mean residence time of 415 years. In the deep sea, R is only 84% corresponding to an age of 1,330 years, which is consistent with the ages of the water masses involved in the thermohaline circulation. All the values reflect steady state conditions and are subject to changes due to natural and anthropogenic effects.

The model in Fig. 13.5.3.1-1 can be expressed in mathematical terms (see Box 13.5.3.3.1). This provides important understanding of the manner in which the carbon cycle determines the properties of the ^{14}C sequestered in tree rings, and other terrestrial archives. A change in production (or some other parameter) is imposed on the model, and the manner in which the change affects the ^{14}C in the atmosphere is computed numerically. Some important examples are given below.

13.5.3.2 The Averaging and Storage Effects of the Carbon Cycle

To illustrate some of the effects of the carbon cycle, Fig. 13.5.3.2-1 presents the transient changes in the model atmosphere following two-step changes in the ^{14}C production rate.

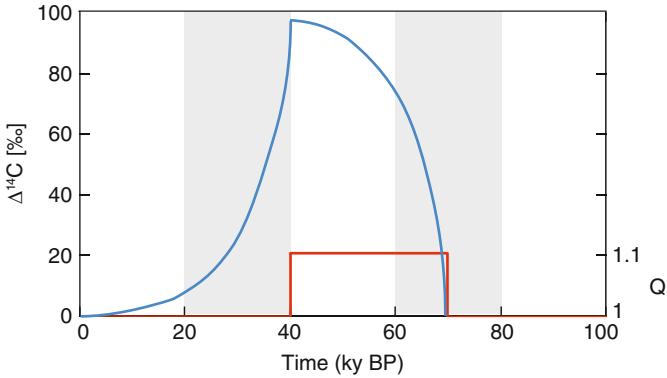


Fig. 13.5.3.2-1 Effect of an abrupt change in the ^{14}C production rate Q by 10% between 70,000 and 40,000 BP. The carbon system reacts exponentially with the characteristic timescale $5,730/\ln(2) = 8,270$ years. In other words, the ^{14}C system needs several half-lives (40,000 years) to reach equilibrium after a production change. Note that the time axis runs from right to left

Table 13.5.3.3-1 Effects of changes in the eddy diffusion constant K and the gas exchange rate between atmosphere and mixed layer k_{am} on the $^{14}\text{C}/^{12}\text{C}$ ratio in the atmosphere, the mixed layer, and the deep sea

Parameter	$^{14}\text{C}/^{12}\text{C}$ atmosphere	$^{14}\text{C}/^{12}\text{C}$ mixed layer	$^{14}\text{C}/^{12}\text{C}$ deep sea
$0.5 \times K$	111	106	84
$1 \times K$	100	95	84
$2 \times K$	95	89	84
$0.5 \times k_{am}$	104	95	84
$1 \times k_{am}$	100	95	84
$2 \times k_{am}$	98	95	84

The present atmospheric ratio is normalized to 100%

Figure 13.5.3.2-1 illustrates how slowly the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio responds to a production change. In this case study, there is a step function increase of 10% in the production rate of ^{14}C at 70000 year BP, and over the next 30,000 years, all the reservoirs slowly approach a new equilibrium which is 10% (100‰) higher than before. Then at 40000 year BP, the ^{14}C production has been reset to the initial value, and the $^{14}\text{C}/^{12}\text{C}$ ratio decreases exponentially with the time constant of $5730/\ln(2)$ and is still approaching the previous level at approximately 0 year. This shows that the ^{14}C system adjusts very slowly to a new equilibrium after a change in the mean production rate. For example, it would take about 10 half-lives (57,000 years) for the ^{14}C concentration to be within 1‰ of its asymptotic value.

Finally, the carbon cycle model has been used to compute the manner in which the atmospheric concentration of ^{14}C would change if the production rate were to vary as a sinusoid of period T and amplitude A . Figure 13.5.3.2-2 plots the attenuation of the amplitude, and the phase lag of the ^{14}C concentration in the atmosphere, versus period. The figure shows that the attenuation and phase lag are both strongly dependent on the period of the production signal. For example, the amplitude of the

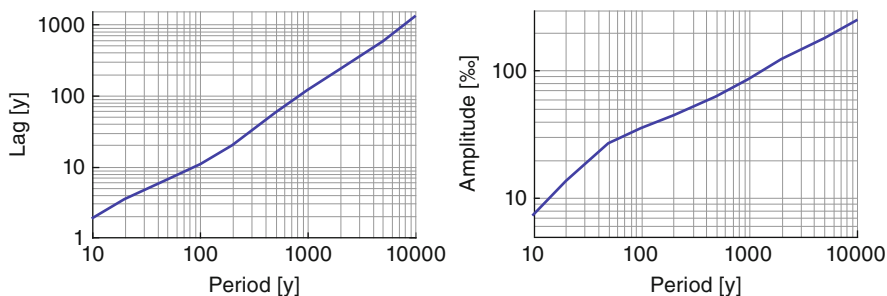


Fig. 13.5.3.2-2 Effect of the carbon cycle on the ^{14}C production signal. Depending on the period of the production signal, its amplitude is attenuated and delayed. Note that the attenuation is given in permil (0.1%)

11-year Schwabe cycle will be attenuated by a factor of ~ 100 so that a 40% 11-year variation (as seen in the ^{10}Be record) will appear as a 4 permil variation in the ^{14}C in tree rings, which makes detection of this cycle in annual ^{14}C data rather difficult. Even more striking is the fact that a production signal at the Hallstatt period ($\sim 2,300$ years) would be attenuated by a factor of 10. The phase lag (left graph) in the first case is about 2 years; that is, the maximum concentration of ^{14}C sequestered in tree rings would occur 2 years after the maximum production rate and as a consequence, the maxima in the ^{10}Be and ^{14}C archives would not be synchronous. The figure shows that this lack of synchronism would occur for all periods, thus the Hallstatt maximum of ^{10}Be would lead that of ^{14}C by ~ 200 years as a consequence of the storage and averaging properties of the carbon cycle.

In summary, the four-box model of the carbon cycle has demonstrated the following points that are vital in our use of the cosmogenic ^{14}C archive:

1. Most of the global ^{14}C ($\sim 95\%$) resides in the oceans and provides a “memory” that has a major influence on the ^{14}C data in the tree ring and other biological archives.
2. As a consequence of this, the atmospheric ^{14}C concentration measured at the time t reflects only to a small extent the production rate at that time, being largely determined by the ^{14}C production history over the previous 20–40 millennia.
3. Further, the carbon cycle results in major changes to the amplitude and phase of the periodic signals we observe in the ^{14}C archive. The reduction in amplitude means that the ^{14}C variation is a factor of 100 less than that in the ^{10}Be record for the Schwabe 11-year cycle; the difference is a factor of 10 for the Hallstatt cycle. The storage effects of the carbon cycle results in the maxima in the ^{10}Be record preceding that in the ^{14}C record by 2 years (Schwabe) and 200 years (Hallstatt).

To date, we have used step changes in the ^{14}C production rate to gain insight into the nature of the ^{14}C record. In practice, the production rate $Q(t)$ is a continuously time varying function and it can be shown that we can write the following convolution integral for the atmospheric ^{14}C concentration, $N_a(t)$, at time t :

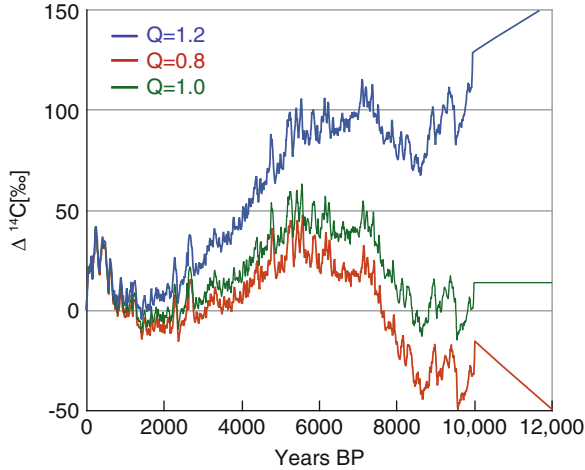


Fig. 13.5.3.2-3 $\Delta^{14}\text{C}$ calculated using the known production rate changes since 10000 BP (from the ^{10}Be record) and three different $Q(t)$ prior to 10000 BP. *Top:* $Q(t) = 1.2$ times the present until 20000 BP, then declining linearly to 1.0 at 10000 BP. *Middle:* $Q(t) = 1.0$ at all times in the past. *Bottom:* $Q(t) = 0.8$ times present until 20000 BP, and increasing linearly to 1.0 at 10000 BP. Note that while the short-term changes, and those due to the changing geomagnetic field are identical in all three curves, the long-term trends are significantly different. Note that the time axis runs from right to left

$$N_a(t) = \int_0^t Q(\tau) e^{-\lambda(t-\tau)} C(\tau) d\tau, \quad (13.5.3.2-1)$$

where

$N_a(t)$: the atmospheric ^{14}C concentration,

$Q(t)$: the production rate,

$C(\tau)$: the effect of the carbon cycle.

To further demonstrate the effects of the carbon cycle, we perform the following experiment using the Oeschger–Siegenthaler box diffusion model (Fig. 13.5.3.1-1). We use the ^{10}Be record as an estimate of $Q(t)$ for the past 10,000 years, and consider three different time variations of $Q(t)$ prior to 10,000 BP. Figure 13.5.3.2-3 presents the results.

It will be noted that all three curves clearly show the Grand Minima throughout the Holocene. They also show the decreasing production rate when the geomagnetic field increased in strength ~5000 year BP, and the increasing production rate as the field decreased after 2000 year BP. However, these well-known variations are superimposed upon large long-term trends that are quite different for the three cases considered. This illustrates that there are large “memory” effects in the ^{14}C data, such as given in Fig. 17.2.1-1. In Sect. 13.5.3.4, we will discuss the way we can largely eliminate these memory effects for applications such as those given in Chaps. 17 and 18.

13.5.3.3 The Sensitivity of the Carbon Cycle to Changing Environments

The previous section has described the manner in which changes in the ^{14}C production signal are recorded in the ^{14}C archives. We now address the allied question of the extent to which changes within the carbon cycle itself will affect the signal in the ^{14}C archive. This is important if, for example, we want to interpret tree ring ^{14}C data in terms of changes in solar activity prior to the commencement of the Holocene. Thus, changing environmental conditions may influence the size of the reservoirs in Fig. 13.5.3.1-1 and the exchange fluxes between them. For example, the vegetation (N_b) may change as a result of a climate change or human land use, the gas exchange between atmosphere and the oceanic mixed layer is influenced by wind strength and sea ice cover, and the eddy diffusion constant K reflects changes in the thermohaline circulation such as are expected to occur when moving from a glacial to an interglacial epoch.

In Box 13.5.3.3.1, it is shown that the change in the $^{14}\text{C}/^{12}\text{C}$ ratio in the atmosphere is given for steady state conditions by

$$\frac{dR_a}{R_a} = -0.11 \frac{dK}{K} - 0.047 \frac{dk_{am}}{k_{am}} - 0.050 \frac{dN_a}{N_a} - 0.009 \frac{dN_b}{N_b} + 0.22 \frac{dh_o}{h_o} + \frac{dQ}{Q}, \quad (13.5.3.3-1)$$

where K , k_{am} , N_a , N_b , h_o , and Q are defined below.

This equation clearly shows that in addition to production changes, the strongest candidates for changes in the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio are the eddy diffusion constant K and changes in the water depth h . To illustrate their relative effects on the different reservoirs, we assume rather large relative changes of a factor of 2. However, before drawing final conclusions, we will briefly discuss potential sources of such changes.

K : (The eddy diffusion constant- Fig. 13.5.3.1-1) Changes of the thermohaline circulation is the main factor influencing K . A 30% reduction in K was assumed for the Younger Dryas period (13,000–12,000) to explain consistently the observed changes in $\Delta^{14}\text{C}$ in tree rings and ^{10}Be in polar ice (Muscheler et al. 2000).

k_{am} : (the exchange coefficient between the atmosphere and the mixed layer) The transfer rate from the atmosphere into the ocean depends on ocean area, which is smaller during glacial epochs (lower sea level, greater extent of sea ice), and on wind speed.

N_a : (the carbon content of the atmosphere, Fig. 13.5.3.1-1) The analysis of air bubbles included in ice cores reveals relatively small changes in CO_2 between 260 and 280 ppm over the past 100,000 years. However, since the beginning of industrialization in the nineteenth century, large amounts of fossil fuel have been burnt, thereby increasing the atmospheric CO_2 content to the present value of about 380 ppm. Since coal, oil, and natural gas are very old, all their ^{14}C has decayed, which leads to a corresponding decrease in the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio. Hans Suess (1955) first noted this effect and it is now called the “Suess effect”. See Fig. 22.2-2.

N_b : (the carbon content of the biosphere) $\delta^{13}\text{C}$ data from sediments suggest a smaller amount of biomass during the last glaciation. However, as a result of the many changes taking place simultaneously, it is difficult to quantify the change in biomass (Zeng 2003).

h_o : (the ocean depth) During glacial times, the sea level was lower by about 120 m, which is a small relative change compared to the mean depth of 3,800 m.

Q : (the ^{14}C production rate) Changes in the solar modulation of the cosmic radiation, and in the geomagnetic field, are the main contributors to variations in the ^{14}C concentration in the atmosphere (see Chap. 10). The potential range of change is about an order of magnitude, from a situation with very high solar activity and a strong geomagnetic field, to very low solar activity and virtually no geomagnetic dipole field. A special example of a strong natural increase in the production rate is the so-called Laschamp event at 41,000 BP when the geomagnetic field almost reversed and its intensity was only about 10% of the present value. According to the model calculation, this event caused a doubling of the production rate (see Chap. 10). The anthropogenic effect due to nuclear bomb tests in the 1950s and 1960s doubled the atmospheric ^{14}C content (see Chap. 12).

Equation 13.5.3.3-1, together with the above estimates, indicates that the greatest changes in the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio between glacial and interglacial epochs (other than production changes) are likely to be due to changes in the eddy diffusivity, followed by changes in the exchange rate between the atmosphere and the mixed layer. The magnitude of these effects for strong changes (factor 0.5, 2 respectively) are given in Table 13.5.3.3-1.

A reduction in K implies a reduction in the thermohaline circulation, and less ^{14}C is pumped down from the atmosphere into the deep sea. For a factor of two decrease, Table 13.5.3.3-1 shows that the $^{14}\text{C}/^{12}\text{C}$ ratio for the atmosphere increases from 100 to 111% of its original value. If, on the contrary, the deep-water formation is enhanced (higher K), more ^{14}C is sucked into the ocean and the atmospheric ratio drops to 95%. The corresponding effects on the mixed layer are smaller and on the deep sea, negligible. A reduction in the gas exchange between atmosphere and mixed layer also leads to an enhancement of ^{14}C in the atmosphere. However, the

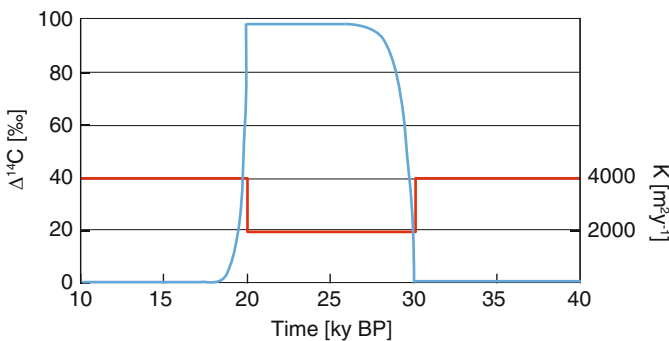


Fig. 13.5.3.3-1 Effect of an abrupt factor of two decrease in the eddy diffusivity between 30,000 and 20,000 BP. The atmospheric $\Delta^{14}\text{C}$ reacts comparatively quickly in both directions

effect is less than 50% of that for the same reduction of K . Clearly, these effects need to be examined when comparing the ^{14}C data obtained during the Holocene, with that obtained during the preceding glacial epoch.

In the preceding section (Figs. 13.5.3.2-1 and 13.5.3.2-3), we have seen that the carbon cycle takes 20–40 millennia to adjust to a change in the ^{14}C production rate. We now investigate the adjustment rate for changes in the parameters in the four-box model shown in Fig. 13.5.3.1-1.

The mathematical box diffusion model was used to compute the consequence of change in the eddy diffusivity K from the present value of $4,000 \text{ m}^2 \text{ year}^{-1}$ to $2,000 \text{ m}^2 \text{ year}^{-1}$ between 30000 and 20000 BP (Fig. 13.5.3.3-1). The result is a comparatively fast change in the atmospheric $\Delta^{14}\text{C}$ by 100‰ (corresponding to 10% in Table 13.5.3.3-1). After 320 years, the $\Delta^{14}\text{C}$ level has risen by 50% and after 470 years by 90%, respectively. That means, the changes in the atmospheric ^{14}C due to changes in the internal parameters of the carbon cycle occur more than a factor of ten faster than the adjustments to production changes.

BOX 13.5.3.3-1 Mathematical Treatment of Sensitivity of the Carbon Cycle to Internal Factors

To estimate the importance of the different parameters of the carbon cycle (Fig. 13.5.3.1-1), we follow the approach of Siegenthaler et al. (Siegenthaler et al. 1980).

Firstly, we make some general rough estimates about the possible ranges of the $^{14}\text{C}/^{12}\text{C}$ ratio. For this purpose, we consider only three reservoirs (atmosphere, biosphere, and ocean) and assume equilibrium between production and decay:

$$Q/\lambda = R_a N_a + R_b N_b + R_o N_o \quad (\text{B13.5.3.3.1-1})$$

with

Q : production rate,

λ : decay constant,

R_x : $^{14}\text{C}/^{12}\text{C}$ ratio in reservoir x ,

N_x : amount of carbon in reservoir x ,

Most of the carbon a,b,o: atmosphere, biosphere, ocean. Most of the carbon is in the ocean and the numbers in Fig. 13.5.3.1-1 yield a mean global R of 85%. Any change in the atmosphere and/or the biosphere will hardly affect the ocean as long as the total amount of carbon is constant. What does this mean for the atmosphere? The minimum possible value of R_a is 84%, assuming complete mixing between atmosphere and ocean. The other extreme would be a complete separation between the atmosphere and ocean, leading to a R_a of 5,780% ($2.4 \times 100 + 1.3 \times 95 + 64.5 \times 84$).

In a next step, we consider the exchange between the atmosphere and the mixed layer of the ocean:

(continued)

$$\lambda R_o N_o = k_{am} N_a R_a - k_{ma} N_m R_m = k_{am} N_a (R_a - R_m) \quad (\text{B13.5.3.3.1-2})$$

assuming steady state conditions, which means $k_{am} N_a = k_{ma} N_m$

According to Fig. 13.5.3.1-1, k_{am} is 1/(7.7 year).

It can be shown (Oeschger et al. 1975) that

$$R_o = R_m \frac{\tanh(h_o \sqrt{\lambda/K})}{h_o \sqrt{\lambda/K}}, \quad (\text{B13.5.3.3.1-3})$$

where h_o is the mean depth of the ocean (3,800 m).

By combining equation (B13.5.3.3.1-1) to (B13.5.3.3.1-3) and considering that R_b approximates R_a , the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio R_a can be expressed as a function of the production rate and the carbon reservoir parameters:

$$R_a = \frac{Q/\lambda}{N_a + N_b + N_o \frac{1}{m + \frac{\lambda N_o}{k_{am} N_a}}} \quad (\text{B13.5.3.3.1-4})$$

with

$$m = \frac{h_o \sqrt{\lambda/K}}{\tanh(h_o \sqrt{\lambda/K})} \quad (\text{B13.5.3.3.1-5})$$

By conserving the total amount of carbon

$N_{\text{tot}} = N_a + N_b + N_o = \text{const}$, N_o can be replaced by $N_{\text{tot}} - N_a - N_b$. Forming the total derivative leads to the following result:

$$\frac{dR_a}{R_a} = -0.11 \frac{dK}{K} - 0.047 \frac{dk_{am}}{k_{am}} - 0.050 \frac{dN_a}{N_a} - 0.009 \frac{dN_b}{N_b} + 0.22 \frac{dh_o}{h_o} + \frac{dQ}{Q} \quad (\text{B13.5.3.3.1-6})$$

13.5.3.4 Computing the ^{14}C Production Rate from the ^{14}C Observations

As discussed in Sect. 13.5.3.2, the temporal changes observed in the tree ring ^{14}C record are strongly modified versions of the ^{14}C production rate itself. Thus, the 11-year variation has been attenuated by a factor of 100; the 2,300-year variation by a factor of 10, while the very long term (>10,000 years) changes suffer relatively little attenuation. As a consequence, the observed ^{14}C record is dominated by the Holocene-long changes, and the production changes due to shorter term solar effects are reduced almost to insignificance.

We therefore need to be able to use the ^{14}C record to compute the time dependence of the ^{14}C production in the past, free of the attenuating effects of the carbon cycle. However, the solution to this problem is not straightforward. There are two methods available to do this, which we now outline in concept. The reader is advised to consult specialist descriptions for further details.

- (a) The first method employs a step by step application of the carbon cycle model in Fig. 13.5.3.2-3. An initial estimate of a change in $Q(t)$ at time t in the past is used to calculate $\Delta^{14}\text{C}(t)_c$. This value is then compared to the measured value of $\Delta^{14}\text{C}(t)_m$. From the difference between the measured and calculated $\Delta^{14}\text{C}$, a correction factor is calculated (using a so-called PID controller) which adjusts $Q(t)$ until $\Delta^{14}\text{C}(t)_c$ and $\Delta^{14}\text{C}(t)_m$ match (Muscheler et al. 2004).

As we have seen in Fig. 13.5.3.2-3, production changes many 1,000 years prior to time t have a significant influence on the observed $\Delta^{14}\text{C}$ record. This therefore introduces a degree of ambiguity into the inverted function $Q(t)$. Thus, different assumptions for the ^{14}C time dependence prior to the commencement of the record will result in somewhat different functions $Q(t)$. Such long-term uncertainties can be largely removed using the ^{10}Be record to estimate the long-term production changes and then adjusting the assumed ^{14}C production history $Q(t)$ prior to the commencement of the record accordingly.

- (b) In the second method, the Fourier transform of the observed $\Delta^{14}\text{C}$ record is first computed (see Box 17.2.1.1 and example in Fig. 17.2.2-1). The decision is then made to limit our attention to production changes with periods less than an upper limit T_U , that is, to frequencies greater than a lower limit $f_L = 1/T_U$. The values of the Fourier transform are set to zero for all $f < f_L$. Then to compensate for the attenuation imposed by the carbon cycle, the Fourier transform is multiplied by the reciprocal of the attenuation function in Fig. 13.5.3.2-2. After correcting the lag, the inverse Fourier transform is then taken, yielding the time variations in the past after elimination of the long-term trends, and correction for the attenuating effects of the carbon cycle (Usoskin and Kromer 2005).

As with the first method, there are some uncertainties in this method. The long-term changes in the estimated production function, $Q(t)$, will depend on the cut-off frequency (f_L) assumed. The sharp cut-off of the Fourier transform at f_L may have undesirable effects, and a more gradual cut-off may be used with resulting changes in $Q(t)$.

Finally, we remind the reader that the ^{14}C record since the mid-nineteenth century has been severely contaminated by anthropogenic factors. The first, the Suess effect (Sect. 13.5.3.3), is the consequence of the steadily increasing addition of CO_2 to the atmosphere from burning fossil fuel. This CO_2 contains no ^{14}C (it having decayed long ago) and results in a major decrease in the $^{14}\text{C}/^{12}\text{C}$ ratio, for which compensation is possible (Tans et al. 1979). The second contamination event occurred with the explosion of nuclear bombs commencing in 1944, and no compensation for this is possible. As a consequence, reconstruction of the ^{14}C production record for the years after 1944 is impossible. Section 6.6 has discussed

the manner in which the paleo-cosmic ray records such as ^{14}C can be inter-calibrated with the modern instrumental record.

13.5.4 The Chlorine Cycle

Most of the chlorine on Earth can be found in the mantle (99.6%), the crust (0.3%), and the oceans (0.1%). The chlorine cycle is not yet in equilibrium. The high solubility in water and the high mobility in soils mean that the rivers continuously carry large amounts of chlorine into the oceans, where it is concentrated by evaporation.

Figure 13.5.4-1 shows a schematic overview of the main chlorine reservoirs and the estimated fluxes between them. For each reservoir, the estimated chlorine content is given. The units are explained in the upper right corner. The Δ value (Tg year^{-1}) indicates whether the incoming and outgoing fluxes are balanced. The arrows reflect the chlorine fluxes (Tg year^{-1}) between the reservoirs. Most of the chlorine is stored in the mantle and the crust. From the mantle, HCl is transferred into the atmosphere by volcanic eruptions. Weathering of Cl -containing minerals is responsible for the input from the crust into the rivers and ultimately the oceans.

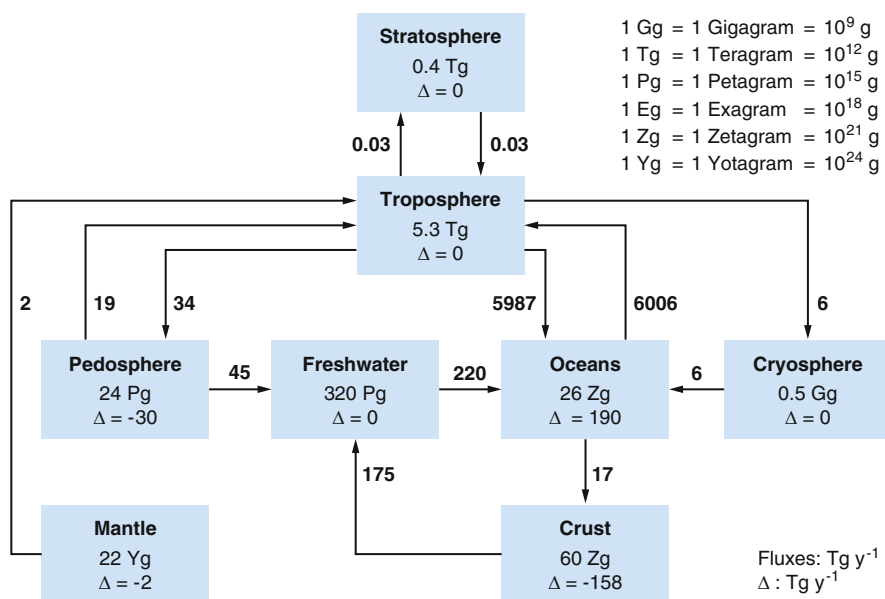


Fig. 13.5.4-1 Schematic overview of the global chlorine cycle. For each reservoir, the estimated total chlorine content and the net flux (Δ value in Tg year^{-1}) are given. The fluxes between the reservoirs are given in Tg year^{-1} . The units are explained in the upper right corner. Note that the chlorine cycle is not in equilibrium. [After (Graedel and Keene 1996)]

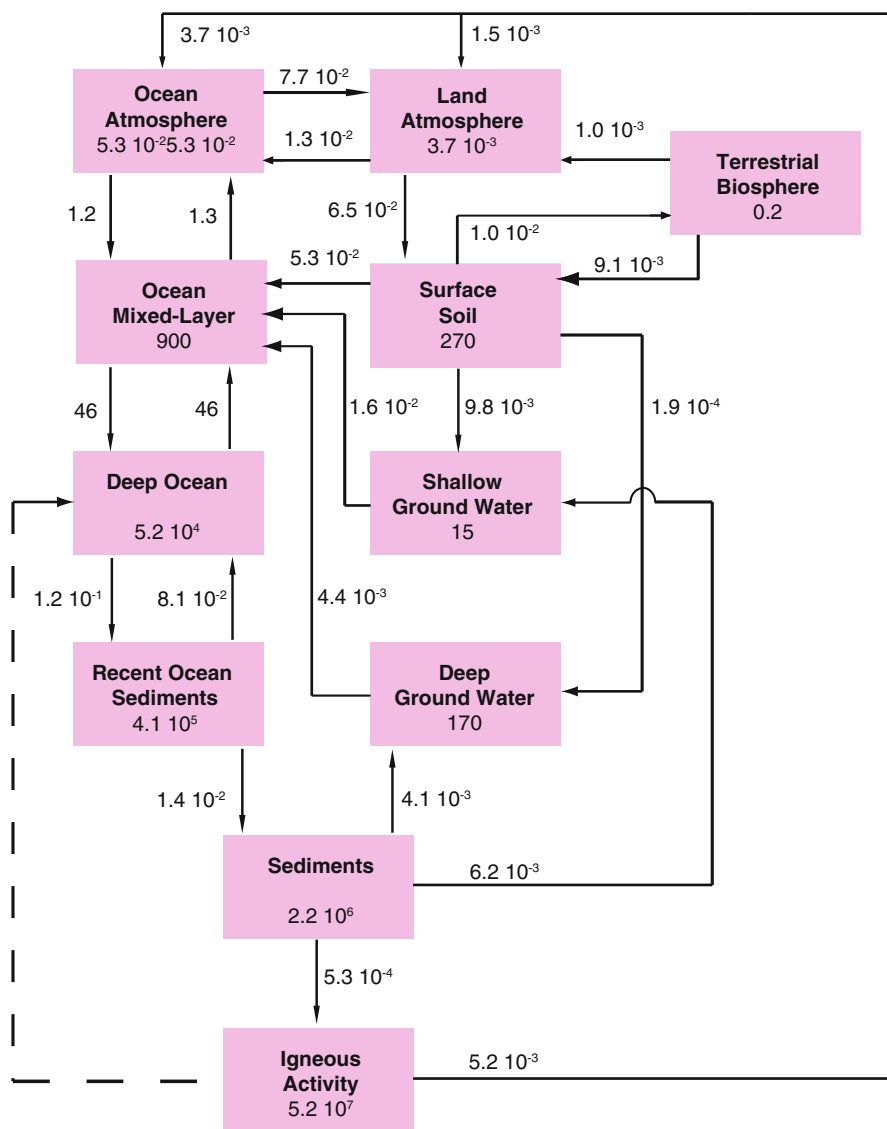


Fig. 13.5.5-1 Iodine cycle after Fabryka-Martin et al. (1985). The inventories are given in grams and the fluxes in grams per year

The third largest reservoir is the ocean with a net annual increase of almost 200 Tg. The greatest loss of Cl from the ocean is to the troposphere. Breaking waves at the sea surface produce sea salt aerosols (sea spray). In addition, a large number of chlorinated organic compounds (e.g. methylchloride) are formed by the biosphere which, however, do not contribute much to the total flux. More than 99% of this flux into the troposphere returns very quickly to the ocean by wet and dry

deposition. In recent times, there is also a non-negligible anthropogenic source of chlorine compounds affecting the atmospheric chemistry (ozone).

To the best of our knowledge, no global ^{36}Cl cycle has been established to date.

It is important to note that most of the ^{36}Cl in the environment after 1950 is due to human activities. Nuclear weapon testing between 1950 and 1963 produced large amounts of ^{36}Cl by neutron activation of ^{35}Cl in the oceans (see Chap. 12). Even larger amounts of ^{36}Cl have been generated inside nuclear reactors by the same reaction on chlorine present in fuel elements, graphite, cooling water, and construction material such as concrete and steel (Hou et al. 2007). Due to its long half-life and high mobility, ^{36}Cl is one of the potentially most critical radionuclides that contributes to radiation doses arising from nuclear waste depositories.

13.5.5 The Iodine Cycle

Figure 13.2-1 presents the ^{129}I cycle based on an estimated global inventory of 1.2×10^{27} ^{129}I atoms (250 kg) (Fabryka-Martin et al. 1985). The inventories in the compartments are given in grams and the fluxes between them in grams per year.

The biogeochemical cycles of other radionuclides such as ^{26}Al are still unknown. Because they are presently of minor importance, we do not discuss them here but refer the reader to specialist publications dealing with the individual elements (Rauch and Pacyna 2009).

References

- Fabryka-Martin J, Bentley H, Elmore D, Airey PL (1985) Natural iodine-129 as an environmental tracer. *Geochim Cosmochim Acta* 49:337–347
- Field CV, Schmidt GA, Koch D, Salyk C (2006) Modeling production and climate-related impacts on Be-10 concentration in ice cores. *J Geophys Res-Atmos* 111:D15107
- Graedel TE, Keene WC (1996) The budget and cycle of Earth's natural chlorine. *Pure Appl Chem* 68(9):1689–1697
- Graedel TE, Crutzen PJ (1993) *Atmospheric change, an earth system approach*. W.H. Freeman and Company, New York, p 446
- Heikkilä U, Beer J, Alfimov V (2008a) Beryllium-10 and beryllium-7 in precipitation in Dubendorf (440 m) and at Jungfraujoch (3580 m), Switzerland (1998–2005). *J Geophys Res-Atmos* 113:D11104
- Heikkilä U, Beer J, Feichter J (2009) Meridional transport and deposition of atmospheric Be-10. *Atmos Chem Phys* 9(2):515–527
- Heikkilä U, Beer J, Jouzel J, Feichter J, Kubik P (2008b) Be-10 measured in a GRIP snow pit and modeled using the ECHAM5-HAM general circulation model. *Geophys Res Lett* 35:L05817
- Heikkilä U (2007) Modeling of the atmospheric transport of the cosmogenic radionuclides ^{10}Be and ^7Be using the ECHAM5-HAM general circulation model. ETH Zürich, Zürich
- Holton JR, Haynes PH, McIntyre ME, Douglass AR, Rood RB, Pfister L (1995) Stratosphere-troposphere exchange. *Rev Geophys* 33(4):403–439

- Hou XL, Ostergaard LF, Nielsen SP (2007) Determination of Cl-36 in nuclear waste from reactor decommissioning. *Anal Chem* 79(8):3126–3134
- Joos F, Gerber S, Prentice IC, Otto-Bliesner BL, Valdes PJ (2004) Transient simulations of Holocene atmospheric carbon dioxide and terrestrial carbon since the Last Glacial Maximum. *Global Biogeochem Cycles* 18:GB2002
- McCracken KG (2004) Geomagnetic and atmospheric effects upon the cosmogenic Be-10 observed in polar ice. *J Geophys Res-Space Phys* 109:A04101
- McHargue LR, Damon PE (1991) The global beryllium 10 cycle. *Rev Geophys* 29(2):141–158
- Muscheler R, Beer J, Wagner G, Finkel RC (2000) Changes in deep-water formation during the Younger Dryas event inferred from Be-10 and C-14 records. *Nature* 408(6812):567–570
- Muscheler R, Beer J, Wagner G, Laj C, Kissel C, Raisbeck GM, Yiou F, Kubik PW (2004) Changes in the carbon cycle during the last deglaciation as indicated by the comparison of Be-10 and C-14 records. *Earth Planet Sci Lett* 219(3–4):325–340
- Oeschger H, Siegenthaler U, Schotterer U, Gugelmann A (1975) Box diffusion-model to study carbon-dioxide exchange in nature. *Tellus* 27(2):168–192
- Rauch JN, Pacyna JM (2009) Earth's global Ag, Al, Cr, Cu, Fe, Ni, Pb, and Zn cycles. *Global Biogeochem Cycles* 23:GB2001
- Scheffel C, Blinov A, Massonet S, Sachsenhauser H, Stan-Sion C, Beer J, Synal HA, Kubik PW, Kaba M, Nolte E (1999) ^{36}Cl in modern atmospheric precipitation. *J Geophys Res* 26(10):1401–1404
- Shepherd TG (2007) Transport in the middle atmosphere. *J Meteorol Soc Jpn* 85B:165–191
- Siegenthaler U, Heimann M, Oeschger H (1980) C-14 variations caused by changes in the global carbon-cycle. *Radiocarbon* 22(2):177–191
- Stuiver M, Polach HA (1977) Discussion reporting of ^{14}C data. *Radiocarbon* 19(3):355–363
- Sturm M, Lotter AF (1995) Lake sediments as environmental archives. *Eawag News*, Eawag, p 44
- Suess HE (1955) Radiocarbon concentration in modern wood. *Science* 122(3192):355–357
- Synal HA, Beer J, Bonani G, Suter M, Wolfli W (1990) Atmospheric transport of bomb-produced Cl-36. *Nucl Instrum Methods Phys Res B* 52(3–4):483–488
- Tans PP, De Jong AFM, Mook WG (1979) Natural atmospheric ^{14}C variation and the Suess effect. *Nature* 280(5725):826–828
- Usoskin IG, Kromer B (2005) Reconstruction of the C-14 production rate from measured relative abundance. *Radiocarbon* 47(1):31–37
- Zeng N (2003) Glacial-interglacial atmospheric CO₂ change – the glacial burial hypothesis. *Adv Atmos Sci* 20(5):677–693

Chapter 14

Archives

14.1 Introduction

The environmental system is complex and dynamic. It involves many different processes, with characteristic time scales ranging from seconds to billions of years, which often interact in a non-linear manner. There is abundant evidence that the present-day environment is markedly different from those prevailing 500, 5,000, and 50,000 years ago, for example. It is clearly pointless to study the present-day environment in isolation; we need to understand the longer-term processes that influenced our world in the past, and will do so in the near future.

And here we have an obvious problem. Enormous progress has been made over the past several decades in analytical methods, and it is now possible to monitor a great number of chemical, biological, and physical parameters with great sensitivity. The space age, and the progress in computer sciences, means that we now measure almost all environmental parameters of interest with high temporal and spatial resolution. Starting in the early nineteenth century, systematic records were initiated of many important parameters which provide a crucial data set for the period prior to the modern era. Before that the data recorded by mankind are limited and sporadic, and we must rely more and more on indirect and proxy data.

One of the strengths of the cosmogenic radionuclides is that their different half-lives provide information about the past on a range of time scales, from decades to millions of years, provided this information has been stored properly somewhere. Now, we all are familiar with the way record keeping has evolved with time. In the middle ages, the monks recorded the occurrence of aurora and the weather in their journals. As science developed, written documents and tables of observations were put sequentially in folders each stamped with a date, often synchronized to within seconds using a nautical chronometer, and later, a timing signal broadcast worldwide by radio [e.g. the WWV call sign broadcast by the United States National Institute of Standards and Technology (NIST)]. Today, most documentary information is stored digitally in computers on hard disks and other devices, again with a

date and time attached, probably derived from a GPS satellite. To be useful, each record must contain (1) an observation and (2) the time it was made.

Fortunately, nature invented its own archives to store environmental information and has also provided its own time scale. In this chapter, we discuss the main archives for cosmogenic radionuclides and how the stored information is retrieved, deciphered, and interpreted. Each archive has its own specific properties regarding the way it stores the information, its temporal and spatial resolution, and the time range it is able to cover. We also discuss the vital issues of the accuracy of the time scale, possible changes and errors in it, and the external time checks that nature has provided us with.

14.2 Intrinsic Properties of the Cosmogenic Radionuclide Archives

To illustrate how nature works in recording environmental information in an archive, we initially consider an example from the modern era: the neutron monitor. In Chap. 6, we saw that neutron monitors were developed to record the intensity of the cosmic rays. Neutrons in the energy range from tens to thousands of MeV are moderated to thermal energies before they reach the interior of a proportional counter filled with BF_3 (or ^3He). The low-energy neutrons are adsorbed by a boron nucleus which then splits into ^7Li and ^4He . The kinetic energy of these two ions is then dissipated by ionizing the gas. The charge impulses created by these ions are then counted, one by one. A man-made neutron monitor is, therefore, a device which records the neutron flux at a specific site by counting online each interaction of a neutron with the gas in the detector.

Nature uses a different approach. In this approach, the atmosphere itself serves as the counting gas. The cosmic rays interact with the nitrogen, oxygen, and other atoms in the atmosphere, yielding atoms of ^7Be , ^{10}Be , ^{14}C , etc. Instead of counting the interaction event between a neutron and the gas, the newly produced cosmogenic radionuclides are used as the detector outputs. The big advantage of nature's approach is that the counting can be done offline centuries or millennia later provided the output nuclei have been properly stored as a function of time. That's when the natural archives come in. As discussed in Chap. 13, most of the cosmogenic radionuclides are removed from the atmosphere and, depending on their geochemical properties, may become deposited on the Earth's surface (e.g. ^{10}Be , ^{36}Cl), or incorporated into the biosphere (^{14}C).

We concentrate on the example of ^{10}Be , which is mainly removed from the atmosphere by either rain or snow. In the latter case, every annual layer of snow in the polar caps contains ^{10}Be atoms that were produced in the previous year or so. Similar to putting sheet by sheet of paper in a folder, nature deposits the snow layer by layer on an ice cap which later turns into ice without any change in the concentration of ^{10}Be . This can go on for hundreds of thousands of years. If we

drill a core from such an ice cap, we can, in principle, count the number of ^{10}Be atoms in a single annual layer at a specific depth. Properly calibrated, this provides a measurement of the cosmic ray intensity at Earth during that year many years ago. In order to know when this layer was formed, however, we need a time scale which gives the relationship between depth and age. That is the topic of Sect. 14.3. For the present, we discuss the main properties of the cosmogenic archive and how it differs from a man-made archive.

Accuracy of the recorded signal. The accuracy of the man-made neutron monitor signal is determined by the count rate, the length of the measuring interval, and the calibration of the device and its stability. Corrections are made for even small changes of the atmospheric thickness above the detector. As a result, modern neutron monitors provide hourly data with standard deviations of 0.3% and better, and excellent long-term stability.

The situation is less favourable in the case of the natural neutron monitor. In a typical case, an annual ice sample contains a few times 10^4 atoms of ^{10}Be in one gram. Since the typical overall efficiency of the detection system (accelerator mass spectrometer, see Chap. 15) is about 10^{-4} for ^{10}Be , a standard deviation of 5–10% is obtained for an annual sample of several hundred grams of ice. Then, as we have seen in Chap. 13, the number of ^{10}Be atoms deposited in the polar caps depends upon the transport within the atmosphere and the efficiency of the scavenging process. Therefore, the data from this natural neutron monitor are subject to climatic changes. Finally, the largest changes in the ^{10}Be concentration occur if the precipitation rate changes on a large scale, as it did during the transition from the last glacial epoch to the Holocene (12,000 years ago). The precipitation rate increased by a factor of 2–3 and although the cosmogenic production rate remained unchanged, the concentration dropped by a corresponding factor (see Fig. 19.4-1). As outlined in Sect. 19.4, the effects of the precipitation rate changes can be minimized by computing the ^{10}Be flux.

Luckily, however, the year-to-year, and longer, changes in the ^{10}Be production rate are larger than those observed by the neutron monitor (Sect. 6.6), allowing most of the effects of the long-term changes in the cosmic ray intensity to be measured with accuracy. Nevertheless, the presence of these climatic and other uncertainties must be borne in mind, and whatever steps possible taken to minimize them (see Chap. 13).

Temporal and spatial resolution. The high counting rates of the man-made neutron monitor means that it can provide a temporal resolution of 5 min, or less, for short-lived phenomena such as the arrival of bursts of cosmic rays from the Sun. Its spatial variability is primarily determined by the spatial variability of the geomagnetic field, which changes slowly over hundreds of kilometres, but can be measured with high accuracy by a mobile neutron monitor, if desired.

The situation is intrinsically different for the natural neutron monitor. The transport and the atmospheric residence times described in Sect. 13.3 dictate that the temporal resolution will be about 1 year. The spatial resolution is also limited by the atmospheric mixing, which is significant especially in the stratosphere where the residence time is a year or more (see Sect. 13.2).

Length of record. The longest continuous neutron monitor records go back to 1951. The data from ionization chambers on the ground and flown on balloons have made it possible to extend the instrumental record back to 1933 (see Chap. 6), although with decreasing resolution and accuracy. But that is the end, and no information about earlier times can be obtained from any man-made device. That is when the natural neutron monitor takes over and shows its strength. It extends the record without any change in accuracy for 10,000 years into the past and provides a unique opportunity to study periods such as the “Grand Minima” when the solar and terrestrial environments were completely different from everything we have experienced during the past 70 years of very high solar activity and low cosmic ray intensity.

Preservation of the archived signal. An important property of an archive is its ability to preserve the stored information over very long times. This is a well-known and quite general problem. For example, written documents decay with time and in the worst case become unreadable. In the modern era, hard disks become corrupted, or storage formats are superseded and become unreadable. In the case of the cosmogenic radionuclides, there is a natural limit set by the half-life of the nuclide of interest. Depending on the power of the analytical method, a cosmogenic nuclide is no longer detectable if its original activity drops below 10% (~ 3 half-lives), 1% (~ 7 half-lives), or 0.1% (~ 10 half-lives) respectively. In addition to this general limitation, each archive has its own specific problems. In the case of the ice which records our natural neutron monitor signal, ^{10}Be shows some mobility on time scales longer than 100,000 year. Such post-depositional processes may disturb the archived signal to some extent and will be discussed for each archive separately (see Sect. 14.4). It is therefore important to consider the quality of an archived signal in detail to eliminate man-made or natural distortions before drawing final conclusions about what happened in the past.

14.3 Time Scales

The value of an archive depends crucially on how well the stored information is dated. Consider the example of the occurrence of climate change in the northern and southern hemispheres. To understand the underlying mechanisms, it may be vitally important to know whether the changes occurred simultaneously, or whether one hemisphere led the other. There are a variety of different dating techniques to choose from: they may be either continuous or discrete, either absolute or relative. The technique chosen may depend on the type of archive being used, and the time interval to be dated. Whenever possible, it is good practice to use more than one technique to improve the accuracy. This also provides a means to estimate the systematic uncertainties.

In Chap. 23, we discuss the application of cosmogenic radionuclides as dating tools. Therefore, we concentrate here on other dating techniques.

Stratigraphic time scales. A very common form of time scale which is available in many different archives is obtained by counting a regular (e.g. annual) structure. The best-known example is that of tree rings which are formed during the growing season and are easy to count. By combining trees which lived at different times but overlap to some extent, long chronologies can be built which cover the last 12,000 years in Europe, to cite one example. The corresponding feature in lakes is the annual laminae called varves, which show a distinct change in composition between summer and winter due to the changing biological activity in the lake during the seasons (Fig. 14.3-1). Another approach makes use of the seasonality of high-resolution records of $\delta^{18}\text{O}$, for example in ice cores, or the presence of partially melted snow each summer. Since $\delta^{18}\text{O}$ reflects the temperature at the time water vapour condenses to water droplets or forms ice crystals, a clear annual signal can be seen (Fig. 14.3-2). An even stronger seasonal signal can be found in H_2O_2 records because H_2O_2 is formed by photochemical reactions, which means there is no production at all during the polar night in wintertime.

At first glance, it would seem that counting layers would be a simple straightforward task. However, there are some pitfalls. For example, trees often do not develop tree rings in tropical areas where there is little seasonal variation. In regions

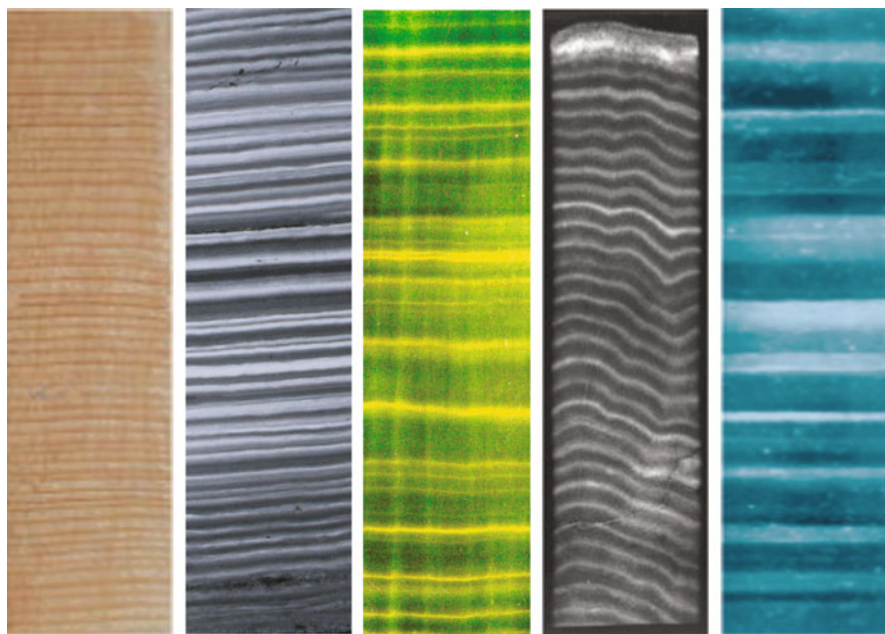


Fig. 14.3-1 Examples of annual structures in archives which can be used to construct a time scale. From left to right: tree rings of Huon pine from Tasmania (Cook et al. 2006); varves from Lake Holzmaar, Germany (Bernd Zolitschka); speleothemes from Rana, Norway (J. Kihle); coral from Papua New Guinea (Sandy Tudhope); ice from GISP2, Greenland (Anthony Gow). (PAGES News, 11, 2003)

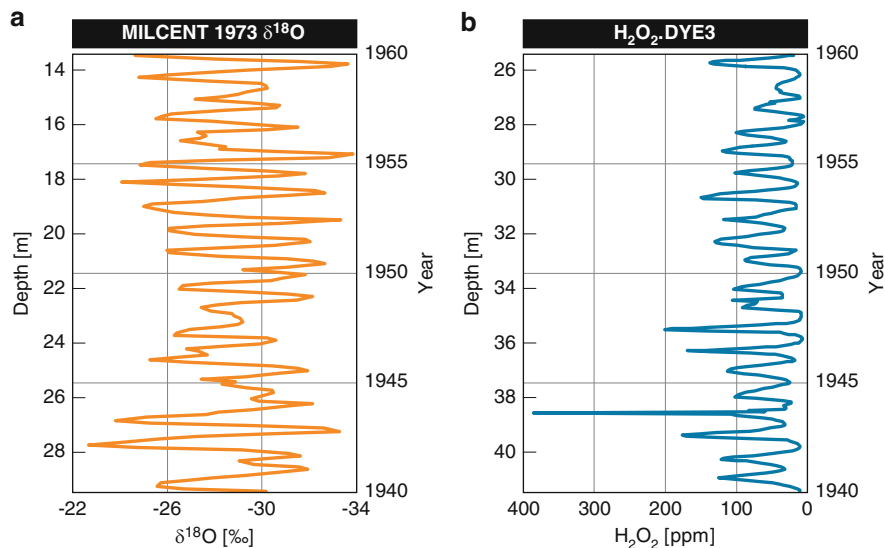


Fig. 14.3-2 Examples of annual signals in isotopic and chemical records. (a) $\delta^{18}\text{O}$ reflects the temperature at the time the precipitation is formed; (b) H_2O_2 is produced by photochemical reactions during the polar summer

with strong seasonality, a tree may skip a year as the result of a hard frost at the beginning of the growing season. Sometimes a tree forms a second ring in the same year. Although such events are rare, small errors can add up when counting over long periods of time. In the case of lake sediments, many different conditions have to be met in order to form varves. Since these conditions are subject to climate change, sediments are hardly ever fully varved and missing varves are more common than one would hope.

The concentration of H_2O_2 produced by photochemical reactions during summer provides the best annual signal following that obtained from tree rings (Fig. 14.3-2). However, H_2O_2 is not very stable and can only be used for the last 200 years or so. While the $\delta^{18}\text{O}$ signal is well conserved, diffusion processes erode its annual peaks, and consequently its application is limited to the past several millennia depending on the accumulation rate and the mean annual temperature. For older ice, dust layers which are usually formed in the spring provide another form of time marker. However, there is no guarantee that there will be one and only one dust peak every year. In general, combining several methods may assist in reducing the errors which can add up significantly when counting long sequences. For this reason, there is a need for discrete time markers that allow the absolute age to be checked.

Time markers. The most common way to establish a time scale for an archive is to use the law of radioactive decay. This topic is discussed in detail in Chap. 23. Time markers provide another way to validate relative time scales. A time marker is typically a clear, short-lived signal in an archive which can be attributed to a well-dated event. A volcanic eruption is a good example; it injects large amounts of ash

and chemical species such as sulphate into the atmosphere, which then spread out and settle down to the surface within a few years. If the eruption is large enough to reach the stratosphere, the fallout can be found worldwide. One of the most famous and historically well-documented volcanic eruptions occurred in the year 1815 in Indonesia. More than 70,000 people were killed when Mount Tambora ejected a volume of $1.6 \times 10^{11} \text{ m}^3$ of material (called “tephra”) into the atmosphere, resulting in time markers throughout the world. The climatic effects were felt globally and 1816 became known as “the year without summer”.

Volcanic eruptions are particularly useful as time markers for ice cores because a wide variety of chemical species are emitted and stored faithfully in the ice. Figure 14.3-3 provides a good example of this. In the case of sediments, ash layers

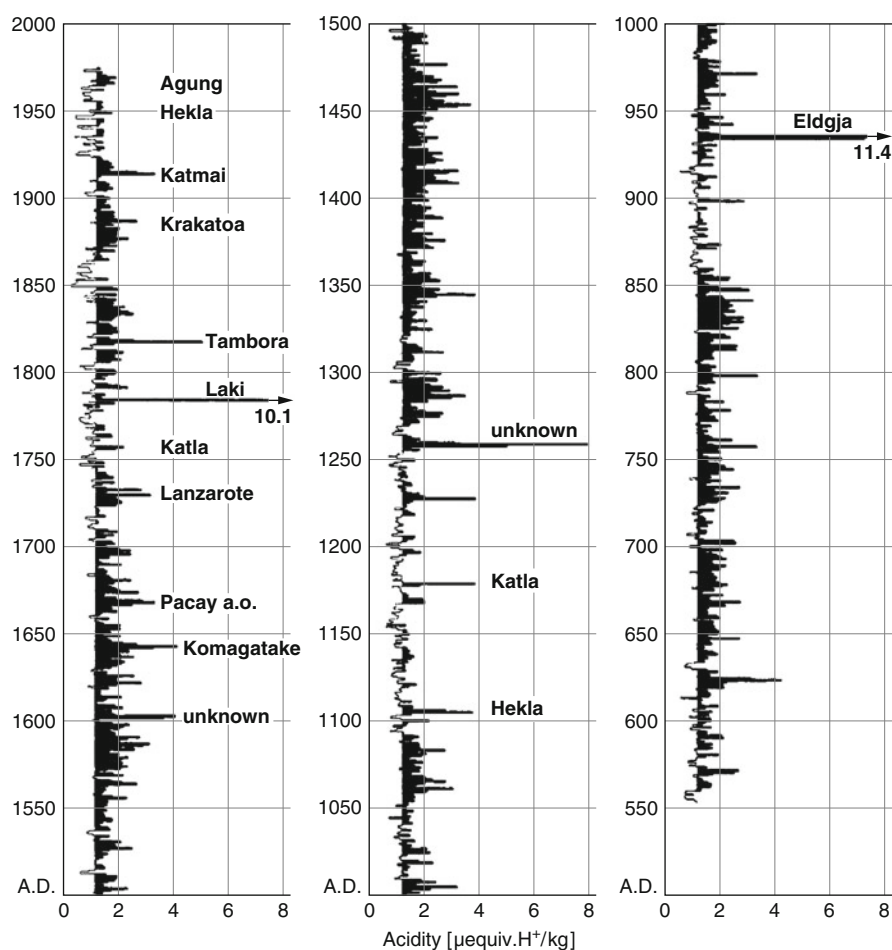


Fig. 14.3-3 Record of historical volcanic eruptions during the last millennium based on acidity peaks in an ice core from the “Crete” drill site in Greenland [after (Hammer et al. 1980)]

can be identified if the eruption has occurred close by, while chemical substances like sulphates are soluble and therefore are not preserved.

Other events that can be used as time markers are the nuclear bomb tests between 1950 and 1963 when most of them came to an end because of the international test ban treaty. They produced large amounts of ^3H , ^{137}Cs , ^{90}Sr , and, ^{36}Cl which entered the stratosphere and were subsequently distributed worldwide. The nuclide ^{137}Cs is particularly suitable as a time marker for sediments because it is chemically reactive and sticks to particles. The Chernobyl accident in 1986 added another ^{137}Cs peak in a large part of the northern hemisphere that is useful for young sediments.

Even a relatively abrupt change in the geomagnetic dipole field can be used by itself as a time marker, or through its effect on the production of cosmogenic radionuclides. The Laschamp geomagnetic excursion that occurred 41,000 years ago is one such example (see Sect. 21.2). In a similar manner, the solar signal in an archive with its own well-defined time scale (e.g. the ^{14}C absolutely dated record) can be used to calibrate the time scale of a ^{10}Be record from an ice core whose time scale is not well known.

Ideal time markers should be unique in the sense that they are global in extent, they cannot be confused with others, and they should have an accurate absolute age attached to them. Unfortunately, this is not always the case. Volcanic eruptions occur all the time, and the recorded amplitude and duration of the signal in a specific archive may depend strongly on the relative position of the volcano, and the local meteorological conditions. As a consequence, there is a danger of mismatching peaks, especially if only relatively short time intervals are considered. For obvious reasons, absolute ages for time markers become a particularly serious problem when going back in time. But even if a time marker is not absolutely dated, it can be very useful to synchronize different archives and to discuss leads and lags of changes in the environmental conditions at different sites.

Finally, it can be useful to model the depth-age relationship of an archive if the formation processes are reasonably well understood. The problem is that such models often need input parameters which are not well known. As discussed later, ice flows as a viscous fluid and the annual layers become much thinner at depth. An ice flow model can provide an estimate of a new depth-age calibration; however, the model will depend strongly on the accumulation rate of the snow. Nevertheless, a model age is helpful in planning a dating strategy. If combined with discrete absolutely dated points, it provides a good means to interpolate between the absolutely dated points.

14.4 Examples of Archives

Table 14.4-1 provides an overview of the most commonly used archives of cosmogenic radionuclides, together with some of their main characteristics. This table is not meant to be complete and will certainly grow in the future. It concentrates on

Table 14.4-1 Overview of the most common archives of cosmogenic radionuclides together with their main properties

Archive	Accum. rate	Temporal resolution	Distribution	Time period	Cosmogenic radionuclides	Main applications
Ice	2–200 cm/year	<1 year	High alt, lat	1 Myr	^{10}Be , ^{26}Al , ^{36}Cl	sol, mag, clim, dat
Lake sediment	0.1–1 cm/year	<1 year	Global	100 kyr	^{10}Be , ^{14}C , ^{32}Si	mag, clim, dat
Deep-sea sediment	0.1–10 cm/kyr	<1 kyr	Oceans	100 Myr	^{10}Be , ^{26}Al	mag, clim, dat
Speleothem	1–10 cm/kyr	1 year	Global	100 Myr	^{14}C	clim
Loess	1–10 cm/kyr	1 year	Mid		latitudes	2 Myr
^{10}Be	clim, dat					
Corals	1–6 cm/y	Days	Tropics	20 Myr	^{14}C	clim
Manganese nodule	<1 cm/Myr	kyr	Pacific	100 Myr	^{10}Be	dat, clim
Peat bogs	10 cm/kyr	>1 year	Global	100 kyr	^{14}C , ^{10}Be	dat
Tree rings	0.1–2 cm/year	<1 year	Global	10 kyr	^{14}C	sol, mag, clim, dat

Abbreviations in the fourth and the last column: *alt* altitude, *lat* latitude, *sol* solar variability, *mag* geomagnetic field reconstruction, *clim* climate research, *dat* dating

the main cosmogenic radionuclides stored in these archives and does not try to cover all the other proxies that may be in the archives. In the following, we add just a few comments about each of them.

Ice. About 80% of the Earth's freshwater is stored as ice, over 99% of it in Antarctica and Greenland. Most of the remaining 20% of the fresh water is underground, with all the rivers and lakes in the world containing less than 1%. Ice is a very special archive in several respects. It is formed from snowflakes which are gradually compacted to form grains called firn. Under the pressure of newly formed snow layers, the compaction continues until the firn turns into ice. The depth at which this transition occurs depends on the accumulation rate and the temperature, and lies between 50 and 100 m corresponding to ages of 100–1,000 years. At this depth, the connection to the atmosphere is cut off and part of the air in the firn is occluded in air bubbles. This property of preserving air samples from the distant past is unique among the archives and provides a powerful means to study the history of the greenhouse gases. Unfortunately, the amount of air in the bubbles is relatively small (100 cm³ per kg ice) and any attempt to measure gaseous cosmogenic radionuclides such as $^{14}\text{CO}_2$, ^{39}Ar , and ^{81}Kr requires large amounts of ice.

Another special feature of ice is the fact that it flows. This feature, well known in mountain glaciers, has important consequences for the depth-age relationship. The increasing thinning of the annual layers with depth leads to an almost logarithmic depth-age relationship (Fig. 14.4-1). Therefore, the longest ice cores from Antarctica (Dome C) represent about one million years. For those cores, each metre of ice at the top represents 250 years and can provide annual resolution. Near the bottom, one metre corresponds to ~1,000 years.

Ice cores are available from the polar regions, and at high altitudes extending all the way to the equator. The advantage of the high-altitude ice cores is that they are

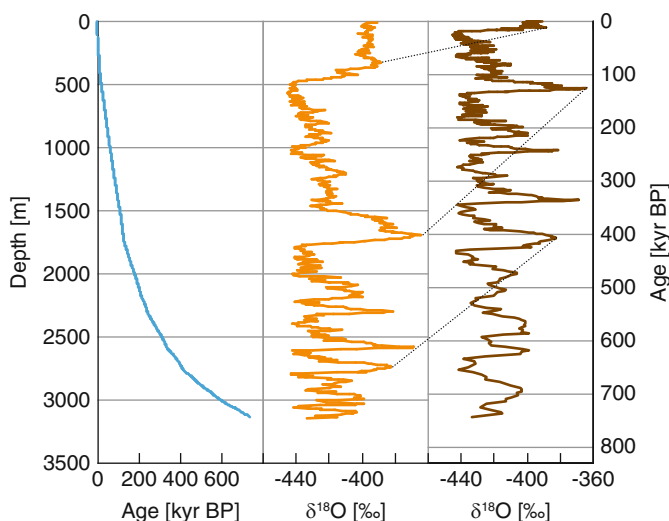


Fig. 14.4-1 Depth-age relationship in the Dome C ice core from Antarctica. Note that the age increases almost logarithmically with depth, resulting in data over a very long period, together with high temporal resolution in the *top part*. This is shown well by the *middle panel*, where the depth scale on the left has been used to plot $\delta^{18}\text{O}$ versus depth. Note the much higher temporal resolution in the first 100,000 years compared to near the bottom of the core. Using the age-depth relationship from the *left panel*, the dependence of $\delta^{18}\text{O}$ on time is given in the *right panel*

closer to the populated areas and provide information on regional effects. However, they are generally restricted to the last few centuries or millennia, and the stratigraphy is often incomplete due to strong winds removing part of the snow after deposition.

At first sight, ice would seem to be the perfect archive because it keeps everything deep-frozen. However, some unexpected phenomena have been observed. For example, part of the deposited ^{36}Cl appears to escape back into the atmosphere causing problems for the $^{10}\text{Be}/^{36}\text{Cl}$ dating method (see Sect. 23.2.3). It seems possible that the H^{36}Cl is released back into the atmosphere during the recrystallization processes in the firn. Another even more unexpected phenomenon is that there are clear indications that ^{10}Be migrates in ice and concentrates on dust particles over time scales of 10^5 years (see Sect. 21.3.2).

Lake sediments. While ice cores are restricted in worldwide distribution, and often hard to access, lake sediments are common throughout the world. They are the result of a variety of complex biogeochemical and physical processes. The particles so formed reflect different environmental conditions within the lake (algal blooms, chemistry) and in the catchment area (erosion, floods, and vegetation). Disturbances of lake sediments can be caused by internal waves, or by earthquakes which lead to landslides on steep slopes. Resuspension by currents within the lake, remobilization of ions under changing chemical conditions, and bioturbation are other sources of disturbances. All these many processes affect the information stored in the

sedimentary layers to some extent. It is easy to find a sedimentary archive, but difficult to find the right one.

As far as cosmogenic radionuclides are concerned, a lake can be considered to be a filter which modifies the input signal (production and atmospheric transport). Thus, the cosmogenic radionuclides signal stored in a sediment consists now of three components: meteoric (direct input from the atmosphere by precipitation), windblown dust, and material from erosion in the catchment area and other changes in the lake system. It is only the insoluble nuclides that are stored in sediments such as ^{10}Be and ^{14}C deposited as carbonate. In most cases, the sediment is strongly influenced by the influx of particles from the catchment area, carrying cosmogenic radionuclides, and often this third component plays a significant if not dominant role as, for example in the case of ^{10}Be . Figure 14.4-2 gives an overview of the various sources of matter which form sediment or are incorporated into it. Rivers can transport large amounts of dissolved and particulate matter into a lake. This input can be amplified when heavy rainfalls cause erosion in the catchment area. In addition, dust and aerosols are washed out from the atmosphere and enter the lake either directly or with the water from the catchment. Finally, biological activity leads to the formation of biogeochemical particles within the lake itself. The sediment layers forming at the bottom of the lake are influenced by bioturbation (mixing by organisms), avalanches on steep slopes, re-suspension, and re-dissolution of deposited material.

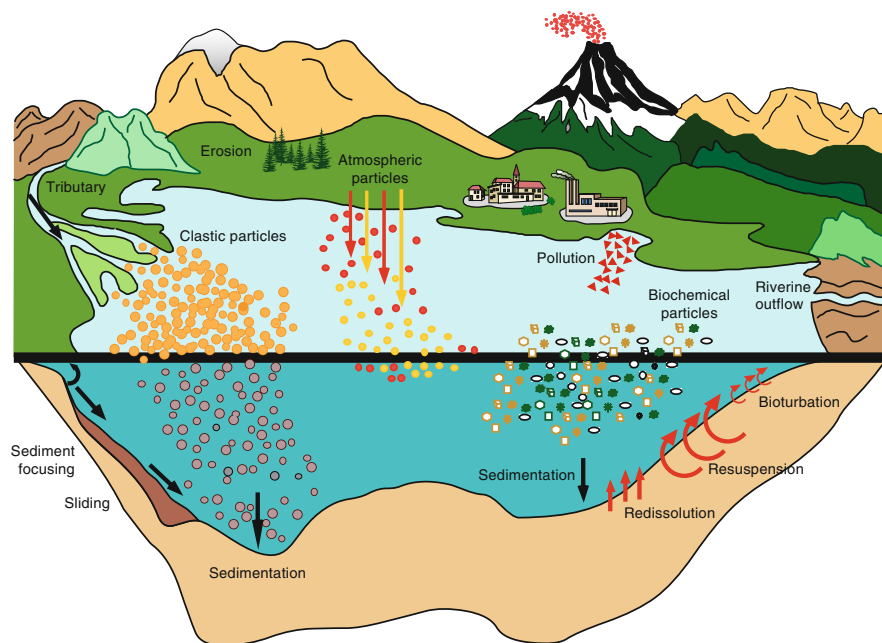


Fig. 14.4-2 Formation processes of lacustrine sediments [after (Sturm and Lotter 1995)]

^{14}C is mainly used to establish the time scale in sedimentary archives. It is important to be aware that the $^{14}\text{C}/^{12}\text{C}$ ratio is generally lower in organic material growing in the lake water as a result of the dissolution of older carbonates, in which some or all of the ^{14}C has suffered radioactive decay. This “hard water” effect typically accounts for timing errors of up to 400 years and can be avoided by selecting land plants that have been incorporated into the sediments.

Ocean sediments. The sea covers two-thirds of the Earth’s surface and consequently sea sediments represent by far the largest archive on Earth. In many respects, sea sediments are similar to lake sediments. However, the input of allochthonous (i.e. foreign) material from rivers is much smaller, and consequently the sedimentation rates are low (see Table 14.4-1). Since the average depth of the oceans is about 4,000 m, drilling deep-sea cores requires large and expensive research vessels. Nevertheless, there have been substantial international drilling programmes and much data are available from this archive (Fig. 14.4-3).

As in the case of lake sediments, the sea can be considered as an additional filter imposed on the input production signal. The majority of the cosmogenic radionuclides are deposited by wet precipitation (Sect. 13.3) into the sea. The most efficient transport of ^{10}Be into the sediment is by particle scavenging, which occurs mainly near the shores where the influx of particles from rivers and dust is the largest. Due to the relatively fast lateral mixing compared to the mean oceanic residence time of 500–1,000 year, the ^{10}Be removed at a site with a large particle

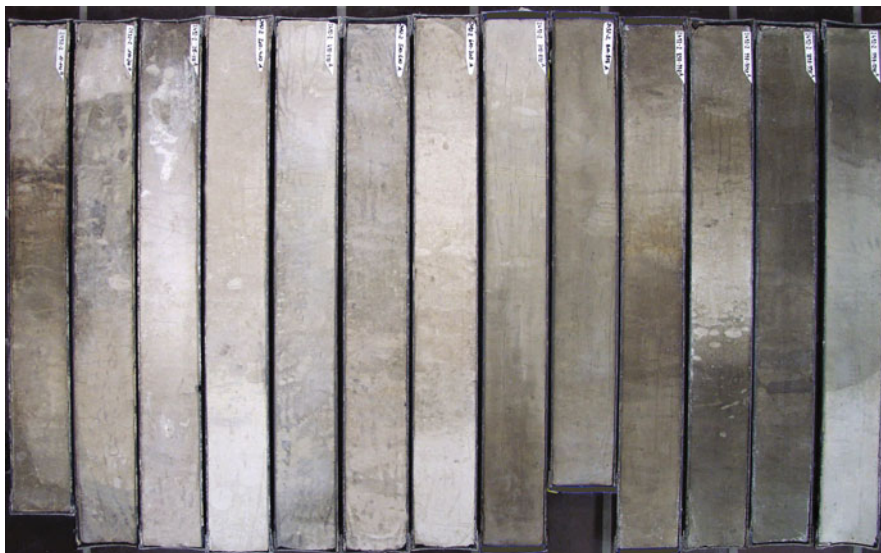


Fig. 14.4-3 Thirteen one-metre-long sections of a sediment core, taken with a gravity corer by the research vessel POLARSTERN in the South Atlantic; light/dark-coloured changes are due to climatic variation during the Quaternary; the core covers about one million years

influx is replaced with ^{10}Be from the surroundings, which may result in very high and inhomogeneous deposition fluxes.

One might think of using ^9Be to correct for this enhanced scavenging effect. Since ^9Be and ^{10}Be react chemically in exactly the same way, the scavenging effect is also the same for both radionuclides. Therefore, one could be tempted to use the $^{10}\text{Be}/^9\text{Be}$ ratio to minimize the focusing effects. However, this approach does not work because the sources of ^9Be and ^{10}Be are completely decoupled from each other. ^{10}Be is produced in the atmosphere, while the ^9Be mainly stems from eroded material brought in by the rivers. Therefore, both are subject to independent changes which affect the $^{10}\text{Be}/^9\text{Be}$ ratio. To correct for this effect, a ^{230}Th normalization is applied. This normalization is based on the following assumption. Ocean water contains a rather constant concentration of soluble salts of ^{238}U . It is the first member of a long decay chain ending with ^{206}Pb (see Fig. 12.2.2-1). On its way to become lead, it forms ^{230}Th . Since ^{230}Th is insoluble, a constant flux of ^{230}Th precipitates into the sediment. Measuring the ^{230}Th production rate in the ocean water and comparing it to the ^{230}Th flux into the sediment provide a correction factor which changes considerably (factor of 2) from point to point (see Sect. 14.7.2).

Sea sediments have a low temporal resolution, but provide information over the longest time period of all the archives (100 Myr) discussed here. The use of radionuclides is largely limited by their half-lives to the past 15 Myr (see Table 14.4-1), the limit being approximately ten times the half-life.

Loess. Loess is a terrestrial sediment formed by the deposition of windblown silt consisting of different minerals (quartz, feldspar, mica, clay, and carbonates) (Fig. 14.4-4). It originates from deserts or mountain areas. Various processes such as weathering and glacial grinding produce fine dust particles that can be picked up by strong winds and transported over thousands of kilometres. Approximately 10% of the continents are covered with loess that can reach a thicknesses of up to 200 m (mainly in China) covering more than 2 Myr. After deposition, the loess changes in various ways depending on the climatic conditions.

The growth of vegetation during warm and wet periods slowly turns loess into soil. Therefore, loess is an archive which provides information on windiness, aridity, and humidity during glacial and interglacial times.

The flux of ^{10}Be into the loess consists of two components, the common meteoric source of ^{10}Be produced in the atmosphere and removed by precipitation, and a component of ^{10}Be attached to the dust forming the loess. The latter is due to ^{10}Be that was deposited from the atmosphere in the source region of the dust, for example in the Gobi desert in China. After becoming attached to the dust grains, the ^{10}Be was then blown by the wind to the loess plateaus (see Sect. 21.3.5).

Speleothems. Speleothems are formed in caves when groundwater comes into contact with air (Fig. 14.4-5). CO_2 escapes to the atmosphere followed by precipitation of CaCO_3 into sequential layers as the pH increases, thereby forming the beautiful stalagmites and stalactites seen in caves the world over. Using recent advances in the U/Th dating techniques, speleothems can now be dated very accurately and provide a very useful new archive for high-resolution climate studies. ^{14}C measurements on appropriately selected stalagmites reveal promising results



Fig. 14.4-4 Loess formations in China. *Upper panel:* Loess gully area of the Shanxi province. *Lower panel:* Loess profile from the Dizhai site showing two dark palaeosol layers corresponding to the previous interglacials around 100 and 200 kyr ago. (Photograph: Maarten A. Prins)

(Hoffmann et al. 2010). The contribution from dissolved CaCO_3 in the soil layer above the cave has to be corrected for as is done in the case of lakes (reservoir effect).

Tree rings. Tree rings are formed during the growing season of trees and contain a large amount of carbon (Fig. 14.4-6). As with all living species, trees are very complex systems which depend on and react to changing environmental conditions in various ways. For example, the growth rate of a tree ring depends on parameters such as temperature, light, water, droughts, nutrients, soil properties, illness, etc. A single frosty night early in the growing season can severely interfere with the growth of a ring even if the conditions during the rest of year are good.

Tree rings represent the classical archive for radiocarbon studies. ^{14}C is well mixed in the atmosphere and consequently, a ^{14}C record measured in one single

Fig. 14.4-5 Speleotheme from Buckeye Creek Cave, West Virginia, USA (photograph: Gregory S Springer, Ohio University)



sequence of tree rings can be used as a reference of the past atmospheric ^{14}C concentration throughout the world. This reference set is used as a calibration curve to correct ^{14}C ages and to turn them into absolute ages. The official calibration curves INTCAL04 (Reimer et al. 2004) and INTCAL09 (Reimer et al. 2009) are primarily based on European oak trees and end (at the time of writing) around 12500 cal year BP because no continuous sequence of overlapping trees has been established so far from the glacial period prior to that time. There is an urgent need to extend this curve for another ~30,000 years. Samples much older than 40,000 years become difficult to measure because 99.7% of the original ^{14}C has decayed by that time.

Corals. Like tree rings, corals are an archive produced by a living organism (Fig. 14.4-7). Corals are produced by small animals (polyps) which live in large colonies mostly in relatively warm, clear tropical water. In their tissue are colourful plant-like cells that use photosynthesis to grow and provide oxygen to the polyp. The whole colony continuously secretes CaCO_3 with the $^{14}\text{C}/^{12}\text{C}$ ratio of the surface layer of the oceanic system, and this is deposited on the coral produced in the previous year.

Coral growth rates vary from about 1 to 6 cm/year. This is illustrated by Fig. 14.4-7, wherein the new coral added in 1 year of growth has been stained with a harmless pink dye. With time, the corals grow to massive and strong reefs which protect many tropical islands from the waves and form very diverse ecosystems in the oceans which are comparable to those in the rain forests on the continents. Coral reefs that can be found today date back to 25 million years, and



Fig. 14.4-6 Tree rings. This is an example which shows the annual rings very clearly. It also illustrates that the annual growth rate is different for different directions, which means that the ring centre is not identical to the centre of the tree. The individual ring width in one direction is variable, reflecting the climatic conditions. These variations are superimposed on a decreasing trend in ring width, which has to be corrected for by a standardization method (photograph: HD Grissino-Mayer)

the topmost 50,000 year can be used as a cosmogenic ^{14}C archive (Druffel et al. 2008). (Section 23.1 shows that the radioactive half-life of ^{14}C limits its use to dating the past $\sim 50,000$ year).

Deep-sea corals are an excellent archive to determine ventilation ages (Frank et al. 2009).

Manganese nodules. Manganese nodules were discovered in 1803. They are rocks that form on the sea floor and consist of concentric layers of manganese and iron hydroxides around a nucleus. Manganese nodules lie on the seabed sediment with a size similar to that of a potato (Fig. 14.4-8). Their composition varies considerably containing up to 70% manganese, some 15% iron, and small quantities of copper, cobalt, zinc, and nickel. The global amount of manganese in nodules is estimated to be thousands of billions of tons and could be of great commercial value. Their growth rate is by far the smallest of all archives and is typically less than 1 cm per million years. The formation of a manganese nodule is a complex mixture of several processes, including metal precipitation from sea water, remobilization of manganese, decomposition of basaltic debris, derivation of metals from hot springs associated with volcanic activity, and metal hydroxide precipitation induced by



Fig. 14.4-7 Coral on artificial hardground from Bermuda, stained for determining growth rate. (Hannes Grobe, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany)

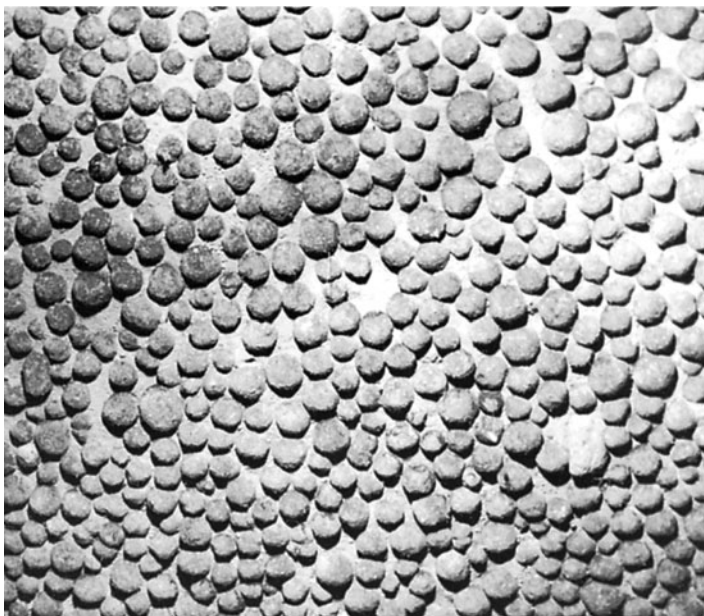


Fig. 14.4-8 Manganese nodules on the southern Pacific Ocean floor (*Encyclopædia Britannica Online*. Web. 29 May. 2011)

microorganisms. Therefore, the composition and growth rate reflect the history of these processes over many millions of years. ^{10}Be is incorporated into the manganese nodule as it grows, and the nodules provide a sequential record of the ^{10}Be content of the oceans over millions of years (Segl et al. 1989).

Peat Bogs. More than 2% of the Earth's land surface is covered with peat. Peat is an accumulation of plant material produced by vegetation such as mosses that has originally grown on the surface. While plant material normally decays rather quickly, this process is impeded in peat-forming systems (mires) due to a lack of oxygen. In most cases, a peat-forming system consists of two layers: a 10–50-cm-thick aerobic upper layer, the acrotelm, with a high hydraulic conductivity and a high rate of decay; and an anaerobic and thicker lower layer, the catotelm, with low hydraulic conductivity and a much slower decay rate. As a consequence of the low hydraulic conductivity, the water table rises to the transition between the acrotelm and the catotelm, provided that it rains enough, as is the case in wetlands. Under steady state conditions, new organic matter is continually formed at the surface, which then travels through the acrotelm, losing about 90% of its mass due to aerobic decay, resulting in the production of CO_2 . The input of carbon into the catotelm is about $50 \text{ g m}^{-2} \text{ year}^{-1}$. There the decay process continues, although at a much slower and anaerobic rate, through the action of methanogenic bacteria (producing CH_4). As a consequence, peat bogs tend to reach a steady state with a zero net accumulation rate and a depth of about 5–10 m because the addition of new matter at the surface is balanced by the losses at all depths. This limits the time span covered by a peat bog to about 50,000 years (i.e. about the dating limit for ^{14}C). Another consequence is that the depth-age relationship is non-linear (Clymo 1984). The organic matter stems largely from sphagnum mosses, which decay and form the peat layers (Fig. 14.4-9) as outlined above.

Peat bogs are ecologically very important because they host a large diversity of plants, insects, and animals, and the ^{14}C in the bogs allows their occurrence to be dated. Peat bogs are also of economic interest: traditionally, peat has been burned in fireplaces instead of wood for thousands of years. More recently, there has been extensive use by gardeners as compost. For example, every year more than 2 million cubic metres of peat are sold for horticultural uses in the UK alone.

Peat bogs are ideal archives for a variety of environmental proxies such as pollen microfossils, plant macrofossils, colorimetric humification, carbon/nitrogen ratios, and others (Langdon et al. 2003; Chambers and Charman 2004). Peat bogs are widely distributed mainly pole ward from the mid-latitudes in both hemispheres.

14.5 Proxies and Surrogates

The information retrieved from a natural archive usually does not provide a direct and undisturbed signal of a single physical parameter. In many, if not most cases, the measured quantity is some correlated quantity that we call a proxy, or surrogate, for the quantity of interest. For example, as discussed in Chaps. 10 and 13, the



Fig. 14.4-9 Peat bog on the Outer Hebrides (photograph: Stu Witmer)

production rates of the cosmogenic radionuclides ^{10}Be and ^{14}C are determined (in part) by the degree of solar activity. Thus, they have been used as proxies for solar activity; while in studies of climate change they have been used as proxies for the solar irradiance (see Sect. 17.5). As discussed in Chaps. 10 and 13, the ^{10}Be and ^{14}C data are influenced by other parameters such as the geomagnetic field, snow precipitation rates, and atmospheric circulation effects, all of which can mislead the solar physicist if their role is not recognized.

The temperatures in the past are of vital interest to many studies, and as in the case of solar activity, there are a number of proxies that we can use. Three such examples are (a) the deviation of the $^{18}\text{O}/^{16}\text{O}$ ratio from a standard value, (b) the thickness of a tree ring, or (c) a pollen profile reflecting the vigour of vegetative growth. In all three examples, temperature plays an important but not unique role. $\delta^{18}\text{O}$ depends not only on the temperature but also on the distance from the source region of the water vapour and the altitude of the site where the precipitation occurs. The thickness of a tree ring depends also on the age and the “health” of the tree and the availability of water. Vegetation vigour and pollen production are also sensitive to water and nutrition supply.

Therefore in these and other proxies, some assumptions (or corrections) must be made in order to estimate the desired quantity from the proxies. In the best case, the physical relationship between the proxy and the parameter of interest is known. The cosmogenic ^{10}Be and ^{14}C are one such case; Chaps. 5 and 7 have described the manner in which the geomagnetic field and the interplanetary medium “modulate” the production of the cosmogenic radionuclides in a manner that can be expressed

in terms of a single parameter, the “modulation function” that can be estimated from the cosmogenic record (Sect. 5.7.3).

In the worst case, the relationship between the proxy and the parameter of interest is not known at all and one has to rely on a statistical relationship (e.g. a regression) between, for example, tree ring width and temperature. The difficulty with this approach is that frequently such a calibration can only be done for relatively recent times for which instrumental data are available (tree ring width versus temperature is one such case). Consequently, it must be remembered that the calibration may not be valid for earlier periods with different climatic conditions.

Finally, we note that one scientists’ proxy may be another investigators’ noise. Thus, Sect. 10.3 has shown that the cosmogenic production rate is modulated by solar activity and on longer time scales by the geomagnetic field intensity. However, this production signal is not directly recorded in an archive. Before being archived, the cosmogenic radionuclides are subject to various transport processes which modify the original production signal (Chap. 13). A solar physicist may call these modifications noise because they disturb the solar proxy that is of interest. On the contrary, for an atmospheric scientist, it may be exactly these modifications that are of interest, because they may assist in the improvement of atmospheric mixing models and lead to a better understanding of stratosphere–troposphere exchange. This is a clear example of the old truth that “somebody’s noise is somebody else’s signal”.

14.6 Properties of Data in the Cosmogenic Archives

14.6.1 Sampling Effects

Finally, having reached an archive with some degree of attenuation and phase lag as specified by the transfer functions (see Sect. 14.6.2), yet more distortion of the cosmogenic signal may result as a consequence of our data taking procedures (McCracken et al. 2004). Thus, for financial or technical reasons, the data samples may correspond to a number of years of deposition, e.g. the commonly used INTCAL04 record of the past 12,000 years of ^{14}C data consists of 5–10 year samples. ^{10}Be records from the high Antarctic plateau frequently use samples that correspond to 6–10 years of deposition as a consequence of the very low snow precipitation rate. The GRIP ^{10}Be data from Greenland are typically 4-year samples for the uppermost part of the core, becoming longer at depth. The length of the samples, or averages that may be calculated later, can have major effects upon the signals of interest in the data.

Consider the case where the data in an archive exhibit a sine wave signal of period T years. Signal processing theory states that if we sample the data with a sample length of S years, then the amplitude of the periodicity in the sampled data relative to the amplitude in the archive is given by

$$\alpha = \frac{\sin X}{X}, \quad (14.6.1-1)$$

where $X = \pi S/T$. This function is plotted in Fig. 14.6.1-1, and it shows that the amplitude of the periodicity is smaller in the sampled data than in the archive. Note that for $S = T$, or $S = 2T$, then $\alpha = 0$. That is, we can either by accident, or by design, sample our data in such a way that some periodicities in the data will be eliminated.

Equation (14.6.1-1) has other implications for the cosmogenic data that are in common use, as we now discuss briefly. To fix ideas, we consider 11-year and 22-year periodicities.

1. Consider the case of ^{10}Be data from Antarctica or Greenland with sampling lengths of $S = 1$ year (e.g. Dye 3) and $S = 4$ years (e.g. the upper Sections of the GRIP core). Figure 14.6.1-1 shows that the attenuations are 0.98 and 0.78 respectively. For longer sampling intervals, with $S = 6$ –8 years (e.g. South Pole), the attenuations are in the range 0.58–0.33; that is, in this case, the sampling process has reduced the signal by a factor of between 2 and 3. From the above, it is clear that three different experimenters, sampling ^{10}Be cores over lengths of 1, 4, and 6 years, would all see an 11-year variation in their sampled data, but the amplitudes would differ in the ratio 1.00; 0.80; and 0.60. Such large differences could cause the unwary to reach totally incorrect conclusions about solar physics or atmospheric processes. Note also the negative sign for the attenuation for $1.0 < S/T < 2.0$ in Fig. 14.6.1-1. This means that the 11-year signal in the sampled data would be out of phase with the signal in the archive. That is, if 16-year “running averages” were taken of the annual ^{10}Be data, there would be a

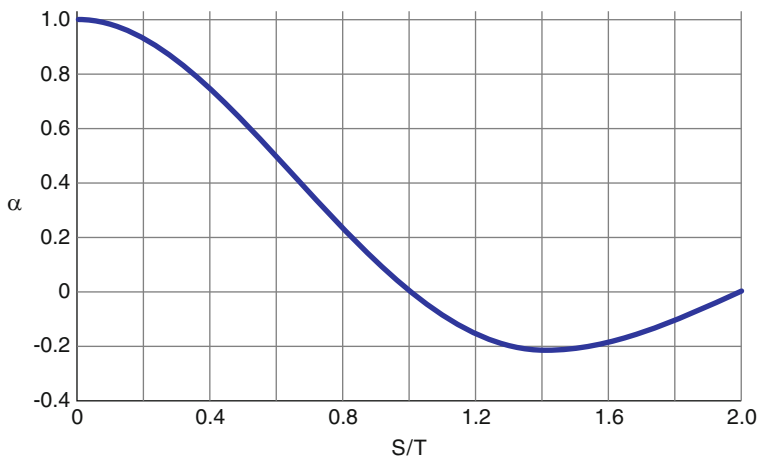


Fig. 14.6.1-1 The dependence of α (amplitude of a periodicity in sampled data relative to amplitude before sampling) on S/T (ratio of sampling interval S to period T)

small 11-year variation in the output (attenuation = 0.2), but it would be out of phase with the signal in the annual data.

2. The widely used INTCAL04 ^{14}C data record consists of 5–10-year averages of tree ring data. Figure 14.6.1-1 indicates that 5-year samples will attenuate an 11-year solar signal by a factor of 0.693 compared to the annual signal in the archive. Figure 13.5.3.2-2 shows that the 11-year production signal has already been attenuated by a factor of 0.009 by the carbon cycle. The overall attenuation is given by the product of successive attenuations so that the total attenuation of the production signal by the carbon cycle, followed by 5-year sampling = $0.009 \times 0.693 = 0.00624$. That is, the 11-year signal, after sampling, is 0.624% of the production signal. In recent times, there has been a 40% 11-year variation in the production signal, so the ^{14}C signal in annual tree ring data is 0.36%, and in the INTCAL04 record 0.25%. In the case of 10-year samples, a similar calculation shows that the attenuation is 0.096, and that the 11-year signal in the annual data has been essentially eliminated in the sampled record.
3. To study the longer-term changes in the cosmogenic data, it is frequently desirable to eliminate the 11- and 22-year solar periodicities. To this end, the data are frequently averaged over 22 years, so $S/T = 2$, and Fig. 14.6.1-1 indicates that the amplitude of both the 11- and 22-year periodicities will be reduced to zero. Figure 7.3-1 is an example of such data.

14.6.2 Transfer Functions

Chapter 13 has discussed the manner in which atmospheric processes modify the cosmogenic signal that is recorded in polar ice and in tree rings. Thus after production, ^{10}Be and ^{14}C follow distinctly different pathways to their archives; ^{10}Be being sequestered in its archive within 1–2 years, while ^{14}C enters the oceans and the biosphere resulting in a very large “memory” effect that extends over thousands of years. That is, the relationships between the input and output (into the archive) are quite different for ^{10}Be and ^{14}C . The characteristics of the transfer processes of ^{10}Be and ^{26}Al to a sea sediment archive are quite different again.

The modification of the production signal by atmospheric transport and other factors can be described mathematically by a “transfer function” (also called the “system response function”). It is a mathematical function that specifies the relationship between the input (production signal) and the output (archived signal). This definition applies only for linear time-invariant (frequently abbreviated LTI) systems, but may also be used with care for systems that are not strictly linear. We have already implicitly used the concept of the transfer function in Sect. 13.5.3.1 when we employed a “three-box model” to compute the characteristics of the carbon cycle in Fig. 13.5.3.2-2. In general, the models used to derive the transfer functions of the cosmogenic radionuclides into their archives are relatively simple two- or three-box models. Despite their simplicity, the transfer functions

provide considerable insight into the information recorded in many archives, as outlined below.

Transfer functions are widely used for precise system calculations by electronic, signal processing, and communication engineers. In cosmogenic studies, they are primarily used to provide us with conceptual guidance regarding what we may expect to see, and what we will not see in the data in our archives. In addition, the attenuation and phase lag for the model approximating the system of interest allow corrections to be made to the data obtained from the archives, and these are vital if we are comparing cosmogenic radionuclides that exhibit different attenuations and lags for the system of interest (e.g. ^{10}Be and ^{14}C in the atmosphere). They have also had limited use in extracting the ^{14}C production signal from the observed archival record as outlined in Sect. 13.5.3.4, and other computational applications can be expected in the future.

In general terms, transfer functions provide us with the following guidance and correction factors for the systems that they describe:

1. Since the models are linear (a mathematical term meaning that all the terms in the transfer function only involve the first power of the input variables), the input periodicities remain unchanged, and no new periodicities are generated. Thus, the 11-year variability in the cosmogenic input signal appears in the output at that same frequency (periodicity) and no harmonics are generated (see Box 17.2.1.1).
2. The transfer function is usually a strong function of period, indicating that some periodicities are strongly attenuated, while the amplitude of other periodicities is little changed in the output to the archive. The transfer function will also show that the phase of a periodic oscillation in the output will suffer a “phase lag” with respect to the input that is a function of period. Both these characteristics are evident in Fig. 13.5.3.2-2. The attenuation and phase lag can be quite different for different cosmogenic radionuclides (e.g. ^{10}Be compared to ^{14}C).
3. The period-dependent attenuation of a transfer function means that it, and the physical system that it represents, can be regarded as a filter. As we will see, the filter characteristics vary greatly from one form of cosmogenic archive to another.
4. The transfer functions allow us to use a periodic signal (Period = T) observed in the archive to estimate the amplitude and phase of that periodicity in the production signal. Thus, if the observed amplitude and phase are A and φ respectively, and the output amplitudes and phase lag of the transfer function for that period are α and $\Delta\varphi$, then the estimated amplitude and phase of the production signal are

$$\text{Production amplitude} = A/\alpha \quad (14.6.2-1)$$

$$\text{Production phase} = \varphi - \Delta\varphi \quad (14.6.2-2)$$

For our purposes, it is enough to understand that the transport of cosmogenic radionuclides from the atmosphere to their respective archives results in a modification of the production signal, which can be described by a transfer function or filter which depends on the basic properties of the system such as size and residence times. Often a complex system such as the atmosphere or the ocean can be very much simplified if the residence times of cosmogenic radionuclides are long enough to average out numerous complex short-term processes. An important point is that although the filtering effect of a system on the production signal can be substantial, it does not mean that the information about the production is lost. As long as the filter properties are known and remain constant (e.g. the transfer function is not affected by climate change), the input signal can be estimated from the output signal in many cases (e.g. the production of ^{14}C despite severe attenuation, Sect. 13.5.3.4). Even if the climatic conditions are not completely constant, the changes in the filter properties may be relatively small, increasing the uncertainty of the estimate of the production signal but not masking it.

In Box 14.6.2.1, we outline the derivation of the transfer function of a simple 1-box system, and discuss its characteristics.

14.6.2.1 BOX Solving Transfer Functions in Box Models

To illustrate the principles in determining a transfer function, we use a very simple example, namely the one-box model (Fig. 14.6.2-1).

Neglecting all the details of the physical transport processes and assuming perfect mixing, the one-box model is completely characterized by the mean residence time T_{res} . The deposition flux D can be described by $D = kN$ with $k = 1/T_{\text{res}}$. If we apply the one-box model for the case of cosmogenic radionuclides, we have to consider an additional sink, the radioactive decay. This leads to the following differential equation:

$$\frac{dN(t)}{dt} = P(t) - k_{\text{sed}}N(t) - \lambda N(t), \quad (\text{B14.6.2-1})$$

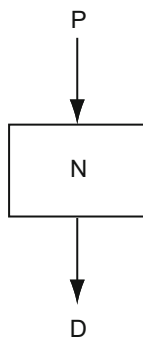


Fig. 14.6.2-1 One-box model with the input function P (for example production) and the output function D (for example deposition)

where $N(t)$ is the number of atoms in the system, $P(t)$ the input function (atoms/time), k_{sed} the loss due to sedimentation (1/time), and λ the decay constant (1/time). k_{sed} and λ can be combined to a single constant k , which leads to the equation

$$N' = P - kN \quad (\text{B14.6.2-2})$$

An elegant way to solve this differential equation is to apply the Laplace transform L and its inverse L^{-1} which is a more general form of the Fourier transform after B14.6.2-2

$$\begin{aligned} sn &= L(P) - kn \\ n &= L(P)/(s + k) \\ N &= L^{-1}\{L(P)/(s + k)\} \end{aligned} \quad (\text{B14.6.2-3})$$

Let us solve the equation for two special cases of the production function P . In the first case, we assume a constant production $P = P_0$. Then, formula (B14.6.2-3) has the form

$$N = L^{-1}\{P/s/(s + k)\},$$

which leads to the well-known solution:

$$N(t) = \frac{P_0}{k}(1 - e^{-kt}) + N_0 e^{-kt} \quad (\text{B14.6.2-4})$$

with $N = N_0$ at the time $t = 0$.

For $t \gg k$, steady state conditions are reached and $P = kN = D$, where $1/k$ is the mean effective residence time.

In the second example, we assume a sinuoidal production rate:

$$P = a * \sin(\omega t)$$

From this, we obtain

$$N = L^{-1}\{aX/(s^2 + \omega^2)/(s + k)\} \quad (\text{B14.6.2-5})$$

The steady state solution is

$$N = \frac{ak \sin(\omega t) - a\omega \cos(\omega t)}{k^2 + \omega^2} + N_0 e^{-kt} \quad (\text{B14.6.2-6})$$

(continued)

Therefore, the steady state sedimentation flux is given by

$$D = k_{\text{sed}} N = k_{\text{sed}} \frac{ak \sin(\omega t) - a\omega \cos(\omega t)}{k^2 + \omega^2} \quad (\text{B14.6.2-7})$$

From this formula, the reduction in amplitude and the phase lag can be calculated as a function of the periodicity T of the production signal ($\omega = 2\pi/T$) and the effective residence time $\tau = 1/k$.

Assuming that λ is much smaller than k_{sed} and can therefore be neglected ($k = k_{\text{sed}}$) as in the case of ^{10}Be and ^{26}Al , we obtain

$$D = \frac{ak^2 \sin(\omega t) - ak\omega \cos(\omega t)}{k^2 + \omega^2} \quad (\text{B14.6.2-8})$$

The results are shown in Fig. 14.6.2-2.

The left panel of Fig. 14.6.2-2 shows the general fact that a sinusoidal signal with a periodicity smaller than the residence time leads to a significant reduction of the amplitude. All curves show consistently an amplitude reduction from 100% to 15.7% for a period equal to the residence time. This means that input (production) changes on time scales shorter than the residence time are in general difficult to detect. The delay in the output signal first increases linearly and then for periods larger than ten times the residence time approaches asymptotically the residence time. For example, a 2,000-year cycle is delayed by 320 years for a residence time of 500 years.

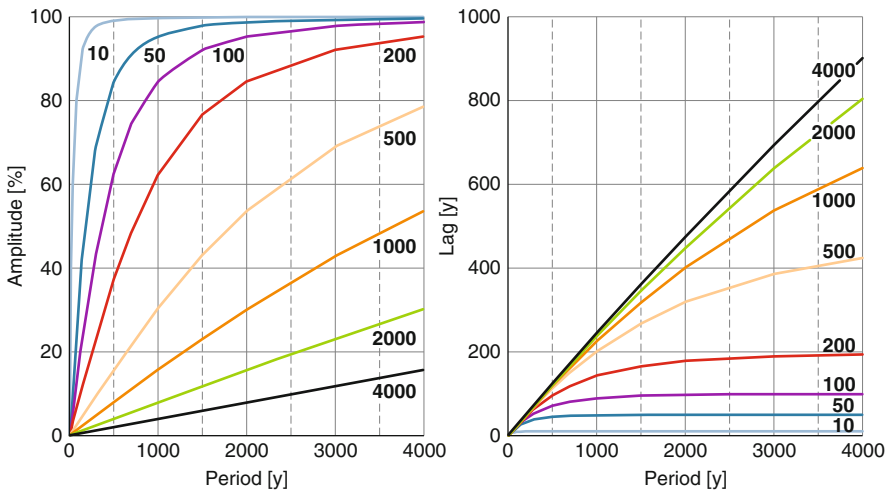


Fig. 14.6.2-2 Attenuation (*left panel*) and lag (*right panel*) for a sinusoidal production signal as a function of the production period for different residence times (given in years)

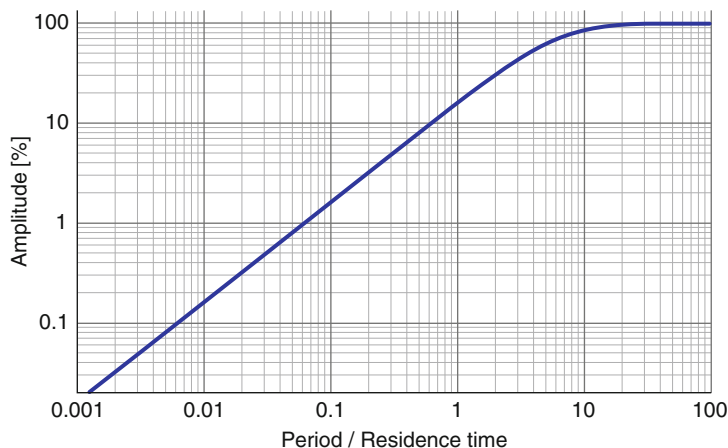


Fig. 14.6.2-3 Relative amplitude of the output signal as a function of the ratio period/residence time

As can be derived easily from formula (B14.6.2-8), the amplitude depends only on the ratio k/ω . This is shown in Fig. 14.6.2-3, where the amplitude is plotted as a function of the ratio of period to residence time. This confirms our result from Fig. 14.6.2-2 that for a residence time equal to the period (x -value of 1), the amplitude is reduced to 15.7%.

If the period is 10% of the residence time, the original signal is attenuated to 1.6%; if the ratio is 10, the amplitude is 84.7%.

Finally, we would like to stress that amplitude and phase change, but not the frequency, as can be seen from formula (B14.6.2-8).

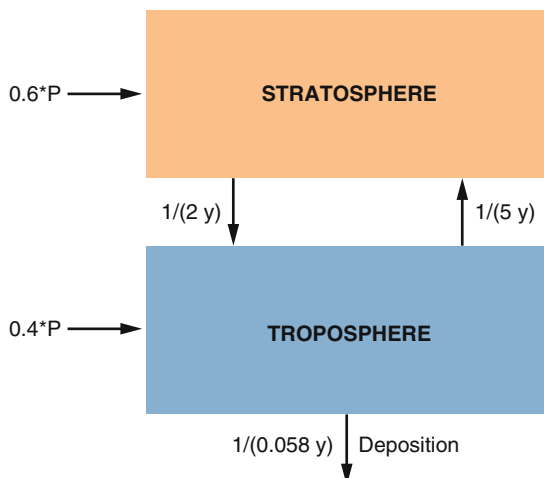
In the case of more complex box models and/or production functions, it may be preferable or necessary to calculate the transfer function numerically.

14.7 Modelled Transfer Functions

14.7.1 ^{10}Be and ^7Be in the Atmosphere

To study the effect of the atmospheric transport of ^7Be and ^{10}Be into the ice archive, we use the concept of box models and design a very simple two-box model consisting of the troposphere and the stratosphere (Fig. 14.7.1-1). We want to stress that such a simplified model will never provide a totally realistic interpretation of a time series measured at a specific site. For this purpose, not only much more realistic but also much more complex general circulation models are required (see Chap. 13). However, a simple model will provide some insights into the

Fig. 14.7.1-1 Two-box model describing the transport of aerosol bound cosmogenic radionuclides (^7Be , ^{10}Be , ^{26}Al , and ^{36}Cl) from the atmosphere where they are produced into the archive ice where they are stored. This model is also used in Sect. 19.2 to investigate stratosphere–troposphere exchange processes



basic effects of, e.g. the atmospheric transport and their dependences on fundamental parameters such as residence times.

We will carry out two experiments using this model. In the first, we simplify the atmosphere even further and make the assumption that all the cosmogenic radionuclides are produced in the stratosphere only (that is, this model approximates the single box model in Box 14.6.2.1). While incorrect, this assumption is helpful in understanding the more realistic experiment in which 60% of the production takes place in the stratosphere and 40% in the troposphere. The mean residence time in the stratosphere is set to 2 years, and in the troposphere to 3 weeks. For the experiments, we use sinusoidal production rates with periods ranging from 1 to 100 years, and compare the amplitude and phase of the deposition flux from the troposphere with the production signal.

The properties of the transfer function are depicted in Fig. 14.7.1-2. The two panels on the left-hand side correspond to the long-lived radionuclides (^{10}Be , ^{26}Al , ^{36}Cl), while the panels on the right-hand side correspond to ^7Be ($T_{1/2} = 53.2$ days).

Firstly, consider the properties of the transfer function for the long-lived radionuclides given in the left-hand panels. The orange curves show the results for the first experiment with production in the stratosphere only. The amplitude for short periods is strongly attenuated (the output is only 8.5% of the input signal for a period of 1 year and 60% for 10 year). For periods longer than 30 years, the output amplitude is greater than 90% of the input, and for periods longer than 50 years, it is greater than 97.5%. In the case of the phase lag (lower panel), the values start with 0.3 year for a period of 1 year and relatively quickly reach an asymptote of ~ 2 years for periods longer than 20 year. That is, periodicities that approximate the mean residence time (2 years in this case) are strongly attenuated, while those that are a factor of ~ 5 and more greater than the mean residence time suffer little attenuation.

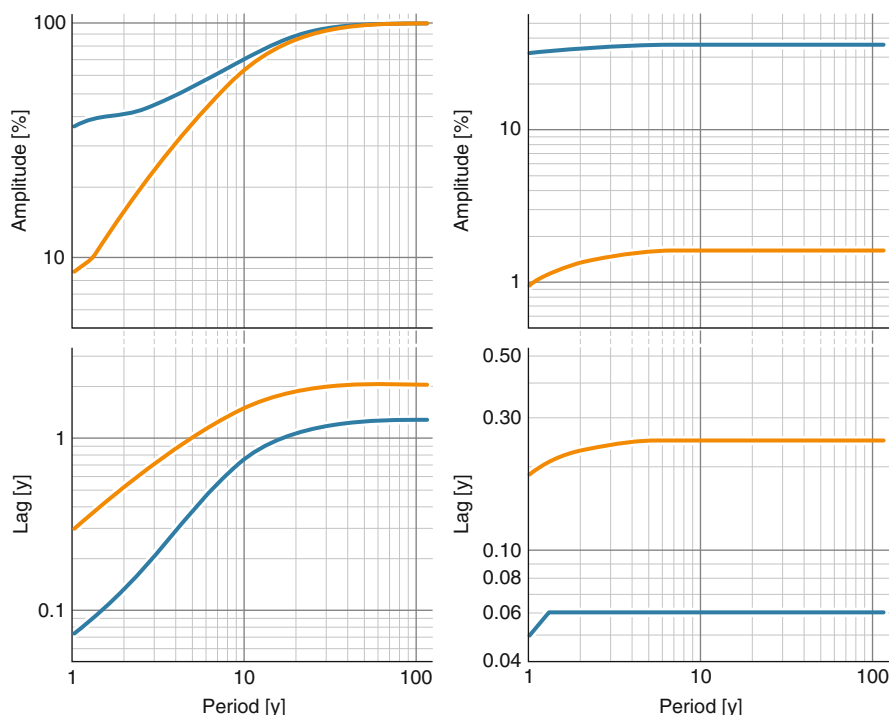


Fig. 14.7.1-2 Transfer functions giving the attenuation and lag for ^{10}Be , ^{26}Al , and ^{36}Cl (left panels) and ^7Be (right panels) caused by atmospheric transport. Orange lines depict stratospheric production alone. Blue lines represent stratospheric and tropospheric production

Now consider the more realistic experiment shown in blue where 60% of the production is in the stratosphere and 40% is in the troposphere. Even for the 1 year period, the amplitude is close to 40%. This is a consequence of the 40% tropospheric component which is deposited in the archive within 3 weeks without significant attenuation. For periods longer than 20 years, the two curves overlap.

In the case of the phase, the lag in the second experiment is systematically shorter, ranging from 0.07 year for a 1-year period to 1.2 years for periods longer than about 30 years. This difference is due to the tropospheric component with its very short residence time.

The right-hand panels show that the ^7Be short half-life of 53.2 days changes the situation dramatically. The amplitude drops to relatively constant values between 1 and 2% if the production is restricted to the stratosphere, because the majority of the ^7Be has decayed before it enters the troposphere. Including tropospheric production raises the amplitude to about 35%. The corresponding lags are 0.25 years (for stratospheric production only) and 3 weeks when tropospheric production is included. This is consistent with the fact that almost all the ^7Be originates in the troposphere.

The main conclusions we can draw from the properties of the transfer functions of this simple model are as follows:

1. ^{10}Be is well suited to study production changes with time scales equal to and longer than 10 years. The output amplitude is still frequency sensitive in the range 70–100% of the original input signal, and the signal is delayed by between 0.8 and 1.5 years; however, corrections for these effects can be made with confidence using (14.6.2-1) and (14.6.2-2).
2. For periodicities longer than 50 years, the attenuation of the ^{10}Be signal is negligible and the lag is ~ 1 year as observed.
3. The ^7Be entering the archive is almost exclusively due to production in the troposphere. The amplitude varies only little between 30 and 35% for periodicities ranging from 1 to 100 years. The lag corresponds to the mean tropospheric residence time (3 weeks in the experiments).
4. The transfer functions of ^{26}Al and ^{36}Cl are expected to be similar to those of ^{10}Be , as given in the left-hand panels of Fig. 14.7.1-2, because all three radionuclides adhere to aerosols and suffer similar transport processes in the atmosphere.

The model in Fig. 14.7.1-1 is used again in Sect. 19.2 to study the properties of stratosphere to troposphere interchange.

14.7.2 ^{10}Be and ^{26}Al in Deep-Sea Sediments

Roughly about two-thirds of the cosmogenic radionuclides removed from the atmosphere are deposited on the ocean surface and ultimately stored in deep-sea sediments if they do not decay (^7Be) or stay in solution (^{36}Cl , ^{129}I). Therefore, the ocean represents a second filter which further modifies the original production signal before it is recorded in the sea sediment archive. As discussed earlier (Sect. 14.4), cosmogenic radionuclides such as ^7Be , ^{10}Be , and ^{26}Al are mainly removed from lake and sea water when they adhere to particulate matter. The particle density is highest along the coast due to input from rivers, aeolian dust blown off the land, and highly bio-productive coastal upwelling. Therefore, as a first approximation, the ocean can be described by a two-box model consisting of an open ocean compartment with a comparatively low particle flux and a coastal compartment with a high particle flux (Fig. 14.7.2-1).

The basic assumption of this model is that the deposition of the cosmogenic radionuclides is controlled by the sedimentation process of particles. Therefore, the input to the model contains two components: the particulate matter S and the cosmogenic radionuclides N . As a result of this empirically confirmed assumption, the nuclide deposition is enhanced in the coastal ocean box. By way of contrast, the input of cosmogenic radionuclides from the atmosphere is assumed to be equally distributed over the whole ocean. Since horizontal mixing of water masses is much faster than vertical mixing, the fast removal of the radionuclides in the coastal box

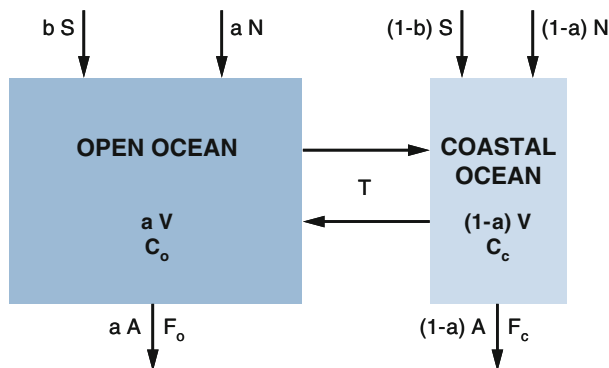


Fig. 14.7.2-1 The ocean is divided into two compartments: the open ocean characterized with a low particle deposition flux and the coastal area with a high flux. From the total area A and volume V , the fraction a is attributed to the open ocean and the fraction $(1-a)$ to the coastal area. S represents particulate matter and N nuclides with concentrations C . S is the total annual particle input in (g/y), N the total annual nuclide input in (atoms/year), C the concentration of radioactive atoms per litre, K the partition coefficient between particulate and dissolved nuclides (l/g), and T the exchange rate in ($\text{m}^3 \text{ year}^{-1}$). Finally, F_o and F_c are the deposition flux densities in ($\text{g cm}^{-2} \text{ year}^{-1}$). After (Rutsch et al. 1995)

leads to a gradient between the open and coastal ocean, resulting in a horizontal flux of nuclides from the open ocean towards the coastal areas. The horizontal water exchange (in both directions) is described by the parameter T in cubic metre per year. The partition of the inputs of nuclides and particles between open and coastal ocean is given by the parameters a and b . Since the open ocean accounts for about 90% of the total oceanic area, a is often set to 0.9. The particle input in the coastal area is much larger than that in the open ocean, and estimates for b range from 0.4 to 0.6.

For steady state conditions (input = output), parameters such as concentrations and deposition fluxes can be derived easily from the model.

To this end, we first calculate the nuclide concentration in the two boxes. The total input of a nuclide into the whole ocean is N . The total output by sedimentation is the sum of the individual outputs from the two boxes $bSKCo + (1-b)SKCc$. Note that KC describes the fraction of the nuclide concentration in the ocean that is attached to particulate matter. Then the output of nuclides transported by particles into the sediment is SKC . Therefore, we obtain

$$N = bSKCo + (1 - b)SKCc \quad (14.7.2-1)$$

If we compile the radionuclide budget for the open-ocean box, the input components are aN from the atmosphere and TCc from the coastal box. The output components are $bSKCo$ into the sediment and TCo into the coastal box. This leads to

$$aN + TCc = bSKCo + TCo \quad (14.7.2-2)$$

These two equations can be solved for C_o and C_c :

$$C_o = \frac{N}{SK} \left(\frac{(1-b)aSK + T}{(1-b)bSK + T} \right) \quad (14.7.2-3)$$

$$C_c = \frac{N}{SK} \left(\frac{(1-a)bSK + T}{(1-b)bSK + T} \right) \quad (14.7.2-4)$$

The radionuclide concentration ratio is then

$$\frac{C_c}{C_o} = \frac{(1-a)bSK + T}{(1-b)aSK + T} \quad (14.7.2-5)$$

For $a = 0.9$ and $b = 0.5$, we obtain

$$\frac{C_c}{C_o} = \frac{0.05 \cdot SK + T}{0.45 \cdot SK + T} \quad (14.7.2-6)$$

If $T \ll SK$, which means that the open ocean and coastal ocean are completely decoupled, then the ratio is 1:9 in agreement with our choice of $a = 0.9$. T is very large: this means the ocean is fully mixed, and the ratio is 1 as is to be expected.

Now we turn to the main question, namely the deposition fluxes into the sediment:

The fluxes from the two boxes into the respective sediments in atoms per unit area and time are given by

$$F_o = \frac{bSKC_o}{aA} \quad (14.7.2-7)$$

$$F_c = \frac{(1-b)SKC_c}{(1-a)A} \quad (14.7.2-8)$$

The ratio of the deposition fluxes is then

$$\frac{F_o}{F_c} = \frac{(1-b)aC_c}{(1-a)bC_o}, \quad (14.7.2-9)$$

and by replacing the concentration ratio using formula (14.7.2-5)

$$\frac{F_o}{F_c} = \frac{1 + \frac{T/SK}{(1-a)b}}{1 + \frac{T/SK}{(1-b)a}} \quad (14.7.2-10)$$

Now we introduce the residence time τ , which is defined as the total number of radionuclide atoms divided by the sedimentation flux (atoms per time) or inventory (atoms cm^{-2}) per flux (atoms $\text{cm}^{-2} \text{y}^{-1}$). As a result of the larger particle flux, the residence time in the coastal box is shorter compared to that in the ocean box. However, we can define a mean oceanic residence time by

$$\tau = VC/(SKC) = V/(SK)$$

Replacing SK in formula (14.7.2-10) by V/τ leads to

$$R = \frac{F_c}{F_o} = \frac{1 + \frac{T\tau/V}{(1-a)b}}{1 + \frac{T\tau/V}{(1-b)a}} \quad (14.7.2-11)$$

Figure 14.7.2-2 shows the dependence of the ratio between the coastal and the open oceanic deposition flux upon the residence time. In one extreme case, when $T = 0$, the ratio between coastal and oceanic deposition flux is 1 because in this case, there is no horizontal mixing and therefore differences in the particle fluxes do not affect the cosmogenic radionuclides. In the other extreme case, if the ocean is perfectly mixed, the ratio R becomes

$$R = \frac{(1-b)a}{(1-a)b} = 9 \quad \text{for } T \rightarrow \text{infinity} \quad (14.7.2-12)$$

This means that the deposition flux in the coastal area can be enhanced up to an order of magnitude for the given parameter a and b . It is interesting to note that the

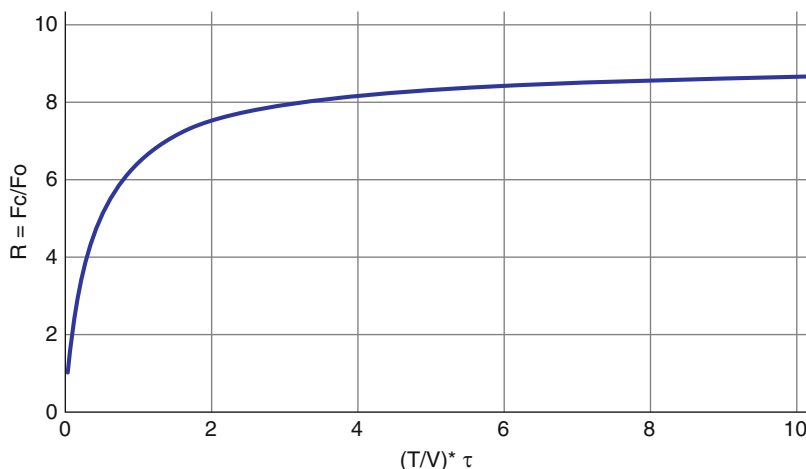


Fig. 14.7.2-2 Ratio between the coastal and the open-ocean sedimentation flux as a function of $(T/V) \cdot \tau$, where T is the water exchange rate ($\text{m}^3 \text{year}^{-1}$) between open and coastal ocean, V is the volume of the ocean, and τ is the mean oceanic residence time; a is set to 0.9 and b to 0.5

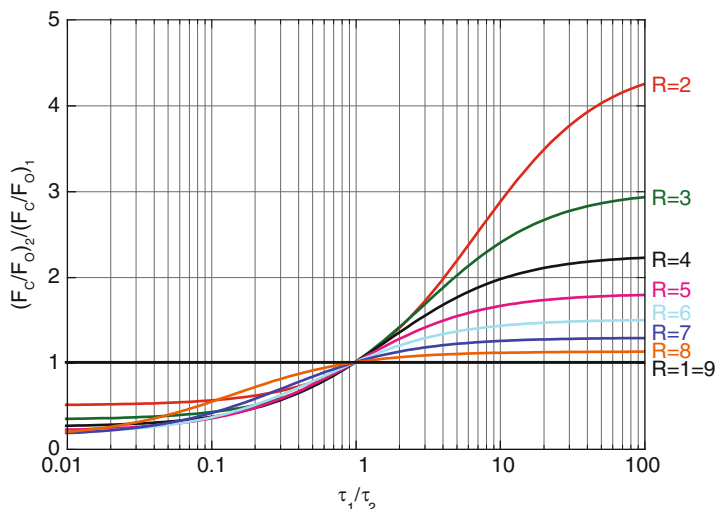


Fig. 14.7.2-3 Ratio of the relative coastal to open-ocean deposition fluxes of two radionuclides with different residence times

ratio R between coastal and open-ocean deposition flux increases very rapidly and has already doubled for $(T/V)\tau < 0.1$.

Finally, we investigate the effect of a change in the mean ocean residence time upon the ratio of coastal to open-ocean deposition flux R for two particle reactive proxies with different residence times. This has practical implications. In Sect. 14.4, we have discussed how a ^{230}Th normalization is used to correct for changes in the ^{10}Be deposition due to particle scavenging in the ocean. As we will see, this correction depends on the ratio of the respective residence times of ^{10}Be (500–1,000 year) and ^{230}Th (7–40 year). For this purpose, we use (14.7.2-11) and obtain

$$\frac{R_2}{R_1} = \frac{\left(1 + \frac{R_1 - 1}{\beta - R_1}\right) \left(1 + \frac{\alpha(R_1 - 1)}{1 - \beta R_1}\right)}{\left(1 + \frac{R_1 - 1}{1 - \beta R_1}\right) \left(1 + \frac{\alpha(R_1 - 1)}{\beta - R_1}\right)} \quad (14.5-20)$$

with $\alpha = \tau_2/\tau_1$ and

$$\beta = \frac{(1 - a)b}{(1 - b)a}$$

Figure 14.7.2-3 shows how the relative coastal to open-ocean deposition fluxes of pairs of radionuclides depends on the ratio of their oceanic residence times $\alpha = \tau_2/\tau_1$. Before we address this practical question, we discuss the result in general terms. If the two residence times of the two radionuclides are equal ($\tau_2/\tau_1 = 1$),

then the relative fluxes are equal as well. For the given conditions ($a = 0.9$; $b = 0.5$) (see Fig. 14.7.2-1), the coastal to open-ocean flux ratio R ranges from 1 to 9 corresponding to a very slow (1) or very fast (9) water exchange rate (see Fig. 14.7.2-2).

As a practical example let us consider ^{10}Be and ^{26}Al . The mean oceanic residence time of ^{10}Be (τ_2) ranges between 500 and 1,000 years, while it is about 30 times shorter (10–30 year) for ^{26}Al (τ_1). If we assume that the coastal sedimentation flux of ^{10}Be is twice as large as the open-ocean flux ($R = 2$), we find from Fig. 14.7.2-3 for $\tau_1/\tau_2 = 0.03$ a value of 0.52. This means (provided everything else is equal) that the ratio of the coastal to open-ocean flux for ^{26}Al is about 1 (0.52×2), which is typical for an element with a short residence time. If however the residence time τ_1 is about an order of magnitude longer than τ_2 as in the case of Zn, we obtain a value of 3 which means that the sedimentation flux of Zn is 6 times (3×2) larger in the coastal zone than in the open ocean.

These examples illustrate that scavenging is strongly enhanced in areas with high particle flux and that it significantly changes the deposition flux from the atmosphere provided that the mean ocean residence time is long enough (a few decades). If the particle flux were constant and known, the model results could be used to correct for the enhanced scavenging effect. However, the particle input depends on many factors and is influenced by the climate. In addition, echo sounding of the sea floor shows that particles do not remain immobile after deposition but are subject to redistribution processes which induce sediment focusing (Fig. 14.7.2-4). Therefore, a better way to correct for local changes in the particle flux is to use an independent proxy for it. Ideally such a proxy will have a constant input flux into the ocean and will be removed from the ocean and transported at the sea floor by particles in the same way as the cosmogenic radionuclides.

^{230}Th is such a constant flux tracer, which almost perfectly fulfils the above requirements. ^{230}Th is a decay product of the primordial radionuclide ^{238}U which has a very long half-life of 4.5×10^9 years. ^{238}U is highly soluble and well mixed in the oceans with a concentration of 3.3 ppb corresponding to a total oceanic activity of 5.4×10^{19} Bq. ^{238}U decays into ^{234}Th , followed by ^{234}Pa , ^{234}U , and ^{230}Th (see Fig. 12.2.2-1). The half-lives of ^{234}Th and ^{234}Pa are short (24.1 days and 6.7 h), and those of ^{234}U and ^{230}Th long (0.25 Gy and 75,400 years). The important point is that ^{230}Th is not soluble and is, therefore, removed from the ocean by particles in a similar way as the cosmogenic radionuclide ^{26}Al . Its residence time is about 50 years. ^{234}Th is not soluble either; however, due to its half-life of 24.1 days, it is much too short lived to be removed before it turns into ^{234}U , which is again soluble. Since the activity ratio of ^{238}U and ^{234}U is in steady state and well known, the production rate of the daughter nuclide ^{230}Th is constant. This means that in every litre of ocean water, 150 ^{230}Th atoms are produced per hour and that at each location, the ^{230}Th deposition flux is proportional to the length of water column above the sediment.

From this, the expected undisturbed sedimentation flux of ^{230}Th can be calculated for each location. In this model, any change due to sediment transport at the

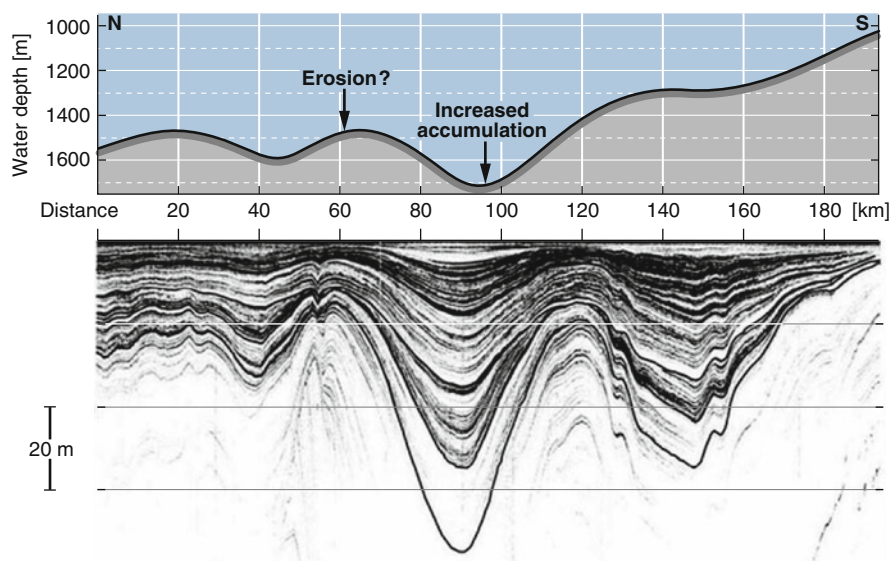


Fig. 14.7.2-4 Echo sounder profile on a section parallel to the coast of Namibia showing evidence for sediment winnowing and focusing [redrawn after (Mollenhauer et al. 2002)]

sea floor leads to a deviation from this value. Therefore, the deviation from the known depositional flux of ^{230}Th can be used to correct for sediment redistribution, which in turn allows calculating vertical accumulation rates of any other sedimentary proxy, e.g. ^{10}Be . In practice, there is an additional effect which has to be taken into account. The constant flux proxy approach does not correct for enhanced scavenging. As described above, a proxy with a comparably long residence time (e.g. ^{10}Be) might be affected significantly by this process. Furthermore, the oceanic residence time of Th may sometimes not be negligible compared to water mass exchange times. Consequently, Th-normalized ^{10}Be fluxes are not corrected for enhanced scavenging and, therefore, do not always reflect the depositional signal from the water column. However, a simple two-box model can be used to estimate this effect and to re-correct the Th-normalized ^{10}Be fluxes for enhanced scavenging of Be over Th. For the details of this correction, we refer to Christl et al. (2003).

References

- Chambers FA, Charman DJ (2004) Holocene environmental change: contributions from the peatland archive. *Holocene* 14(1):1–6
- Christl M, Strobl C, Mangini A (2003) Beryllium-10 in deep-sea sediments: a tracer for the Earth's magnetic field intensity during the last 200,000 years. *Quat Sci Rev* 22(5–7):725–739
- Clymo RS (1984) The limits to peat bog growth. *Philos Trans R Soc Lond B Biol Sci* 303(1117):605–654

- Cook ER, Buckley BM, Palmer JG, Fenwick P, Peterson MJ, Boswijk G, Fowler A (2006) Millennia-long tree-ring records from Tasmania and New Zealand: a basis for modelling climate variability and forcing, past, present and future. *J Quat Sci* 21(7):689–699
- Druffel ERM, Robinson LF, Griffin S, Halley RB, Southon JR, Adkins JF (2008) Low reservoir ages for the surface ocean from mid-Holocene Florida corals. *Paleoceanography* 23(2)
- Frank N et al (2009) The Holocene occurrence of cold water corals in the NE Atlantic: implications for coral carbonate mound evolution. *Mar Geol* 266(1–4):129–142
- Hammer CU, Clausen HB, Dansgaard W (1980) Greenland ice sheet evidence of post-glacial volcanism and its climatic impact. *Nature* 288(5788):230–235
- Hoffmann DL, Beck JW, Richards DA, Smart PL, Singarayer JS, Ketchmark T, Hawkesworth CJ (2010) Towards radiocarbon calibration beyond 28 ka using speleothems from the Bahamas. *Earth Planet Sci Lett* 289(1–2):1–10
- Langdon PG, Barber KE, Hughes PDM (2003) A 7500-year peat-based palaeoclimatic reconstruction and evidence for an 1100-year cyclicity in bog surface wetness from Temple Hill Moss, Pentland Hills, southeast Scotland. *Quat Sci Rev* 22(2–4):259–274
- McCracken KG, McDonald FB, Beer J, Raisbeck G, Yiou F (2004) A phenomenological study of the long-term cosmic ray modulation, 850–1958 AD. *J Geophys Res-Space Phys* 109:12103
- Mollenhauer G, Schneider RR, Muller PJ, Spiess V, Wefer G (2002) Glacial/interglacial variability in the Benguela upwelling system: spatial distribution and budgets of organic carbon accumulation. *Global Biogeochem Cycles* 16(4):1134
- Reimer PJ et al (2004) IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–1058
- Reimer P et al (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* 51(4):1111–1150
- Rutsch HJ, Mangini A, Bonani G, Dittrichhannen B, Kubik PW, Suter M, Segl M (1995) Be-10 And Ba concentrations in West-African sediments trace productivity in the past. *Earth Planet Sci Lett* 133(1–2):129–143
- Segl M, Mangini A, Beer J, Bonani G, Suter M, Wölfl W (1989) Growth rate variations of manganese nodules and crusts induced by paleoceanographic events. *Paleoceanography* 4(5):511–530
- Sturm M, Lotter AF (1995) Lake sediments as environmental archives. *Eawag News*, Eawag, p 44

Chapter 15

Detection

15.1 Introduction

The importance of analytical methods in the field of cosmogenic radionuclides cannot be over emphasized. The practical application of all the theory outlined herein depends strongly on accurate, efficient, and relatively cheap analytical measurements. The classical way to detect cosmogenic radionuclides was to take advantage of their radioactive properties and detect either the actual decay or the decay product. A good example of these two different approaches is provided by ^3H which decays into ^3He through beta decay with a half-life of 12.3 years. In the first approach, the ionization pulse produced by the beta particle is counted using liquid scintillation counters (LSC) or proportional counters. In the second approach, the water containing the ^3H is sealed in a container after pumping out all the dissolved gases. After a few weeks, enough ^3He has accumulated in the container to be measured with a mass spectrometer.

The obvious limitation to the use of radioactive decay is that for long half-lives the counting rates are very low, making this technique very time consuming and inefficient. Furthermore, the concentration of a radioactive nuclide with a short half-life may be very low requiring very great sensitivity. For example, the natural $^{14}\text{C}/^{12}\text{C}$ ratio in the atmosphere is about 10^{-12} . Measuring such a low ratio is comparable to finding a single person in a crowd 100 times the Earth's population. Detection of the proverbial needle in the haystack is a comparatively simple task.

The situation changed dramatically in the late 1970s when accelerator mass spectrometry (AMS) was invented. This method counts the ions of the radionuclide of interest without waiting for them to decay, and allows determination of isotopic ratios to high precision. This technique immediately increased the sensitivity for the measurement of long-lived cosmogenic radionuclides by 5–6 orders of magnitude. This is comparable to the difference between observations of the stars in the night sky by naked eye to using the 5 m Mount Palomar telescope. Suddenly, the door opened for a wide range of new applications and provided great improvements in existing applications. This process of improvement is still going on and it is the

hope of the authors that this book will help and inspire readers to invent their own improvements to this immensely powerful technique.

15.2 Low-Level Decay Counting

As outlined in Sect. 15.1, this technique uses electronic detectors to count the rate, dN/dt , at which the nucleus of interest decays. The decay rate or activity A is given by

$$\frac{dN(t)}{dt} = A = -\lambda N(t) \quad (15.2-1)$$

where λ is the decay constant, which equals $\ln 2/T_{1/2} = 0.693/T_{1/2}$, where $T_{1/2}$ is the half-life. Thus, having measured A , the number N of atoms of interest in the sample is: $N = A/\lambda$.

Integrating the (15.2-1), the number of nuclei of interest in the sample will decrease in the following manner

$$N(t) = N_0 e^{-\lambda t} \quad (15.2-2)$$

and for small half-lives (large decay constants), this means that successive measurements will need to be corrected for the time between each measurement.

In general, the detector will use a pulse height analyser that identifies the pulses due to the decay of interest, and rejects all others. However, there are several practical problems:

1. There may be atoms in the equipment around the sample that decay giving pulses of the same height as those of interest (“equipment background”)
2. The cosmic radiation may generate nuclear reactions in or near the detector that masquerade as the pulses of interest (“cosmic ray background”)
3. There are some decay reactions that result in a wide range of pulse heights from the detector, exacerbating effects (1) and (2) above.

Let us illustrate the third problem using the example of ^{10}Be and ^7Be . ^{10}Be decays to ^{10}B by turning a neutron into a proton (β^- decay) releasing a total energy of 556 keV. Besides the problem of the long half-life (1.387 Myr) which makes decay counting very inefficient, there is the additional difficulty that the 556 keV is shared between the emitted electron and an antineutrino. Therefore, the electron has no single characteristic energy which would enable us to unambiguously identify it as the result of a ^{10}Be decay. Furthermore, there may be other radionuclides in the sample which undergo β^- decay, thereby emitting electrons in the same energy range which cannot be distinguished from the ones stemming

from ^{10}Be . The only solution to this problem is to chemically purify the sample in order to remove as much of the other radionuclides as possible.

The situation is more favourable in the case of ^7Be which decays to ^7Li by electron capture. First of all, the half-life of 53.2 days is short enough to make decay counting an efficient method. Secondly, 10.5% of the resulting ^7Li nuclei are in an excited state, and subsequently decay releasing a 476.6 keV gamma ray. This energy is very well defined and specific to the ^7Be decay, making it easy to discriminate against other decays occurring at the same time in the sample.

As outlined above, a detector has a “background” comprising three components. (1) cosmic rays interacting with the detector or its surroundings, (2) radioactive nuclides in the close environment (e.g., U/Th in the concrete of the building, Radon in the air), and (3) radionuclides within the detector system itself. These components result in a “background rate” that must be determined with accuracy and subtracted from the sample counting rate. There are serious statistical limitations to this procedure which requires that careful attention be given to minimizing the background as far as possible to get the best results.

As illustrated in Fig. 15.2-1, there are two main defences against the cosmic-ray-induced radiation. The first is the use of mass to absorb the cosmic radiation; sometimes the systems are installed deep underground for this reason. The second is the use of an anticoincidence shield – a large particle detector (usually a proportional counter or plastic scintillator) that detects the cosmic ray secondaries (e.g., muons which have a longer mean free path than protons and neutrons as discussed in Sect. 12.2) before they reach the sample detection system. Electronic

Fig. 15.2-1 Ge-detector for gamma spectroscopy. The detector is shielded by 10 cm lead (*top cylinder*) against environmental radioactivity and by an anticoincidence proportional counter on *top* (the *flat copper box*) against cosmic ray muons. The Dewar vessel below the detector is filled with liquid nitrogen to reduce the electronic noise in the solid-state (Ge) gamma ray detector (collection J. Beer)



systems then “veto” any count from the sample detector that is coincident with the detection of a cosmic-ray-induced event by the anticoincidence shield. The system shown in Fig. 15.2-1 has been optimized to detect radionuclides that emit gamma rays (e.g., ^7Be) by using a liquid-nitrogen-cooled solid-state detector that detects the gamma rays with high-energy resolution. It is shielded by 10 cm of lead (top cylinder) against the background from the environment. On top of it is a flat proportional counter which detects cosmic-ray-induced muons which can interact in the counting system producing spurious gamma rays. It detects about $100 \text{ muons m}^{-2} \text{ s}^{-1}$ and reduces the overall background by a factor of 2.

Despite the use of shielding, anticoincidence systems and careful elimination of spurious radioactivity, a detection system still has a background counting rate C_b when the sample is not in place. Then, if C_s is the counting rate with the sample in place, the counting rate due to the sample alone is $C_s - C_b$. In this case, the statistical variances add, so the standard deviation of the result (square root of the sum of variances) is increased by the statistical fluctuations in both the background rate and the sample rate. Consequently, for samples with low radionuclide concentrations, the statistical fluctuations in the background dominate the final uncertainty.

Today, low-level decay counting for cosmogenic radionuclides is mainly used for ^3H , ^7Be , ^{14}C , and ^{44}Ti . ^3H and ^{14}C decay by β^- decay while ^7Be and ^{44}Ti have characteristic gamma lines which make detection with a high resolution Germanium detector easy. ^3H and ^{14}C are both frequently measured with proportional counters using a counting gas that contains the radionuclide of interest (e.g., CO_2 or CH_4 in the case of ^{14}C). Alternatively, the ^{14}C may be synthesized into benzene and counted in a liquid scintillation counter. These counting techniques and the corresponding sample-preparation procedures have been improved over the years to a very high degree of performance which guarantees high precision results and justifies their use for specific samples in spite of their need for large sample sizes and long measuring times. In practice, measuring times of 1 week are common, and in some cases (e.g., ^{44}Ti) it may be much longer. As a consequence, throughput is usually small for all low-level counting systems. Nowadays, radionuclides with half-lives longer than a thousand years such as ^{14}C are increasingly measured using AMS as described in the following section.

15.3 Accelerator Mass Spectrometry

As discussed in the introduction, the invention of the AMS was a major breakthrough in the detection of long-lived cosmogenic radionuclides. Instead of waiting for it to decay, the nuclide itself is counted. However, counting the absolute numbers of a nuclide like ^{14}C or ^{10}Be in a sample is extremely difficult for technical reasons discussed later. Further, only a relatively small fraction of the atoms in the sample can be counted. In practice, it is much easier to measure an isotopic ratio in a sample such as $^{14}\text{C}/^{12}\text{C}$ or $^{10}\text{Be}/^9\text{Be}$. Mass spectrometers, which can do this very precisely, had been used for a very long time prior to the 1970s; so why did it take

so long to apply them to cosmogenic radionuclides? The simple answer is that the isotopic ratios in the case of cosmogenic radionuclides are much smaller, in the range of 10^{-8} – 10^{-15} , compared to no less than 10^{-6} in conventional mass spectrometry. To go to such extremely low ratios requires a much stronger suppression of interfering masses due to molecules (e.g., between ^{13}CH and ^{14}C) or isobars (different elements of the same atomic weight; e.g., ^{14}C and ^{14}N). Ingenious ideas were attempted to overcome these problems; e.g., Anbar's (1979) attempts to develop a technique to measure ^{14}C . To avoid interference by the isobar ^{14}N , which is the most common isotope in the atmosphere, Anbar tried to measure the ratio $^{14}\text{C}^{15}\text{N}/^{12}\text{C}^{15}\text{N}$. The attempt failed due to ^{29}Si sputtered from the anode, showing that a new approach was required to solve the background problems.

That new approach arose initially from an investigation in 1939 when Alvarez and Cornog (Alvarez and Cornog 1939) used an accelerator to demonstrate that ^3He was stable. From this observation, they drew the conclusion that ^3H was unstable and therefore radioactive. Inspired by this work, Muller (1977) recognized the potential of accelerators to suppress backgrounds and proposed to use cyclotrons for the detection of ^3H , ^{14}C and ^{10}Be .

At the same time, nuclear physicists experimenting with negative ion beams made an interesting discovery: nitrogen does not form stable negative ions (Bennett et al. 1977; Nelson et al. 1977; Purser et al. 1977). Immediately they recognized the importance of their discovery: there is no isobaric background due to ^{14}N if negative ions are used instead of positive. The key was to develop a mass spectrometer based on a tandem accelerator which is operated with negative ions. The molecular background is still a problem however; e.g., ^{13}CH and $^{12}\text{CH}_2$ will interfere with the measurement of ^{14}C . This problem can be solved by making use of high energies as outlined below. The basic concept of a Tandem accelerator mass spectrometry system is shown in Fig. 15.3-1.

The solid sample material is inserted into an ion source and exposed to a caesium beam which sputters the sample material producing negative ions. The ions are accelerated by 30–40 kV to form a beam. They then enter the first mass filter where the negative ions with the correct mass are selected and injected into the Tandem

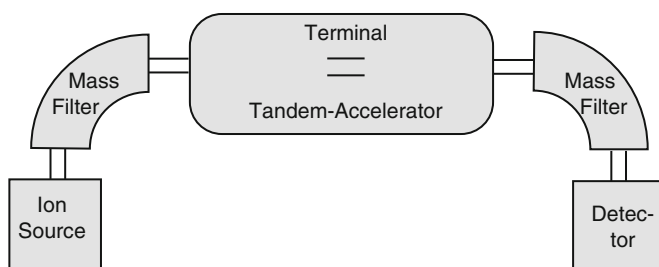


Fig. 15.3-1 Basic concept of an AMS-system consisting of an ion source, a Tandem accelerator with a stripping unit, a detector and two mass filters. The negative ions are accelerated to high energies to break up the molecules in the stripper and to separate the isobars by measuring the specific energy loss of the positive ions reaching the detector

accelerator where they are accelerated twice (that's why it is called "tandem"). By applying a high positive voltage of several million volts to the terminal half way through the big tank in the diagram, the negative ions are accelerated for the first time. As they pass through a thin foil (or gas) inside the positive terminal the electrons are stripped from them, yielding positive ions which are accelerated a second time towards the right-hand end of the Tandem in the diagram. The stripping region serves another vital purpose in that all the molecules are broken up into their atomic components as they pass through and form positive ions with different charges. Following passage through the second mass filter, each atom is identified in the detector by measuring its specific energy loss (dE/dx). According to the non-relativistic formulation of the Bethe formula (10.2.2-1), the energy loss is proportional to the mass times the square of the charge on the ion divided by the kinetic energy ($dE/dx \propto mz^2/E$). In this way, isobars can be suppressed very efficiently. For example, while the isobars ^{10}Be and ^{10}B may both pass through the preceding mass and charge filters and then enter the final detector, a fully ionized boron atom ($z = 5$) will have a $5^2/4^2 = 1.56$ times greater rate of energy loss than that of a Beryllium atom ($z = 4$), which allows us to distinguish between them.

In practice, AMS systems are more complicated than outlined above, and have additional filters using several different techniques (magnetic, electrostatic, or time of flight). Figure 15.3-2 shows a simplified schematic of the AMS system at the ETH Zurich used for the measurement of the $^{10}\text{Be}/^9\text{Be}$ ratio (Suter 2004). The high current ion source (HCS), where negative ions of BeO are formed, is on the left-hand side. In the first 90° magnet, a high voltage switches 50 times per second allowing alternate passage of $^{10}\text{BeO}^-$ and $^9\text{BeO}^-$. The ions are focussed by electrostatic lenses and injected into the 6 MV Tandem accelerator (^{10}BeO 99% of the time; ^9BeO 1%), and the $^9\text{BeO}^-$ current with mass 25 is measured in a Faraday cup (LE-cup). In the Tandem the ions are accelerated to the central terminal, where the molecules are broken up and electrons stripped from the ions.

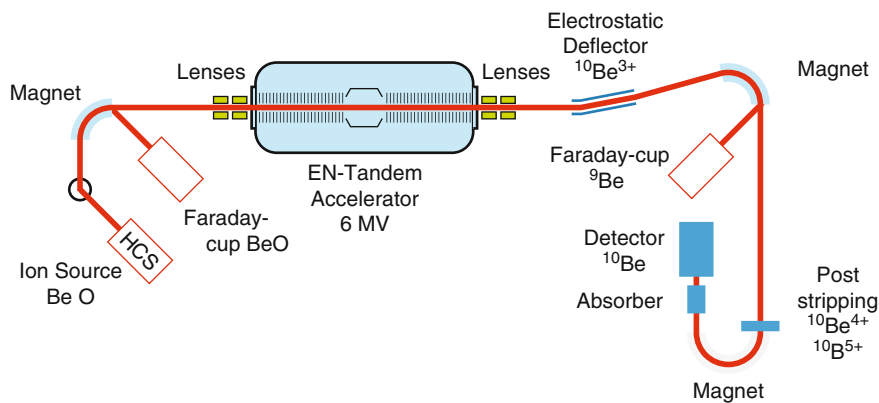


Fig. 15.3-2 Simplified schematic view of the ETH Zurich AMS system based on the 6 MV Tandem accelerator (Courtesy M. Christl)

At this point, the ^{10}Be ions have charge states between 0 and 4+, and they will acquire markedly different energies in the second acceleration stage. The ions then pass through an electrostatic deflector tuned to select the charge state $^{10}\text{Be}^{3+}$ corresponding to an energy of about 18.6 MeV. Since ^9Be is lighter than ^{10}Be , it is deflected more strongly in the following magnet and enters the high-energy Faraday cup (HE-cup) where its intensity is measured. This allows the transmission of the system to be derived from the ratio between the HE-cup and the LE-cup. The ^{10}Be ions continue their journey and cross a stripper foil at the entrance of an additional magnet (180°), which removes the remaining electrons from the ions. Although this post-acceleration stripping reduces the efficiency of the system somewhat, it has the great advantage that it reduces the interference by the isobar ^{10}B . Thus ^{10}B , which has almost exactly the same mass as ^{10}Be , will be stripped to charge +5 while ^{10}Be is restricted to +4. Therefore, the final 180° magnet separates ^{10}B from ^{10}Be very efficiently. The final separation takes place in the detector using the dE/dx technique. By switching rapidly between the two isotopes the $^{10}\text{Be}/^9\text{Be}$ ratio is almost continuously measured and any drifts in the beam intensity leaving the ion source are removed. Figure 15.3-3 is a partial view of the ETH AMS facility, and it is evident that it is even more complex than outlined above.

In the early days, general-purpose nuclear physics tandems (and even cyclotrons) were used as AMS with acceleration voltages up to 12 MV. It is present-day practice for AMS to be dedicated special purpose systems, and there is a clear tendency to use acceleration voltages as low as 0.2 MV depending on the nuclide to be measured. The very low energy machines can only measure ^{14}C .

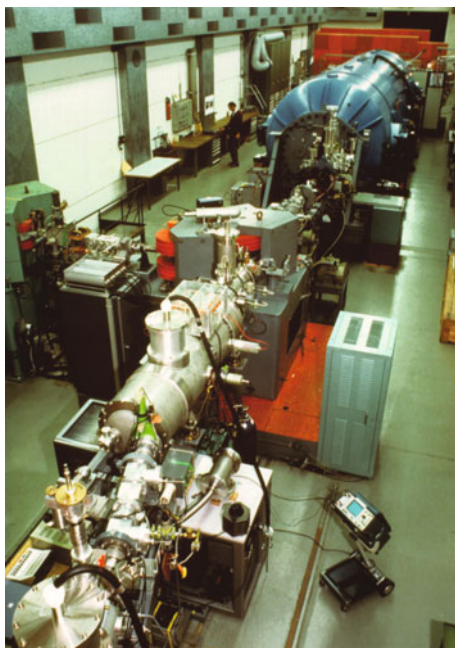


Fig. 15.3-3 Partial view of the AMS system of ETH Zürich consisting of the 6 MV Tandem accelerator (*blue tank*) and the 40 keV electrostatic deflector. The ion source at the far end and the detector at the front end are both invisible. The total length of the AMS system is 30 m (Collection J. Beer)

Worldwide, there are more than 60 AMS systems in operation and the number is growing steadily.

The principle of sample preparation for the AMS technique is outlined in Box 15.3.1.

15.3.1 BOX Sample Preparation for AMS

A mass spectrometer measures isotopic ratios such as $^{14}\text{C}/^{12}\text{C}$ or $^{10}\text{Be}/^9\text{Be}$. ^{14}C which is produced in the atmosphere forms $^{14}\text{CO}_2$ and enters the carbon cycle (see Sect. 13.5.3), where it becomes well mixed with the stable isotopes ^{12}C and ^{13}C . Therefore, the preparation of a carbon sample for AMS is rather straightforward. The carbon is extracted from a sample, in most cases by combusting it and collecting the CO_2 . The final steps of the sample preparation depend on the specific design of the ion source. Often, a solid sample of elemental carbon is pressed into a holder which is then mounted in the ion source. Fractionation effects that may occur during the sample-preparation process can be corrected for by measuring the $^{13}\text{C}/^{12}\text{C}$ ratio (Box 23.2.1.1).

However, most other cosmogenic radionuclides are not globally mixed with their respective stable isotopes. For example, there is no ^9Be in the atmosphere except in dust grains. This means that ^{10}Be extracted from an ice sheet can only be analysed with AMS if some ^9Be is added. This is usually done by putting a small but well-known amount (typically 0.1–2 mg) of ^9Be (called a “spike” or a “carrier”) into the sample at the beginning of the sample preparation. When the ice melts, ^{10}Be and ^9Be mix homogeneously, and the $^{10}\text{Be}/^9\text{Be}$ ratio is fixed and cannot change even if some material is lost during subsequent chemical steps. These steps are designed to extract Be from the sample (water, sediment, soil, and rock) in as pure a state as possible using standard chemical procedures (ion exchange, precipitation, etc.). In the case of ^{10}Be , special attention is paid to reduce Boron as much as possible because it interferes as an isobar with ^{10}Be in the AMS system. After the measurement, the number of ^{10}Be atoms in the sample can be calculated from the measured ratio R and the known amount of ^9Be added: $^{10}\text{Be} = R \times ^9\text{Be}$. The AMS systems are designed to measure extremely small isotopic ratios and their relative changes with high precision. However, the absolute ratios (accuracy) depend on the tuning of the system and have to be calibrated. This is done by measuring standards with well-known absolute ratios. The normalization of the measured data to an internationally accepted standard also guarantees that the data from different AMS facilities can be compared. By alternately measuring samples and standards, any drift in the AMS system can be detected and corrected for.

The instrumental background is another source of uncertainty. Even if a sample contains no cosmogenic radionuclides, the detector will have a small counting rate. There are many different sources contributing to the background. Examples are incomplete suppression of isobars and scattering of isotopes on gas atoms within the AMS system. There is also a danger of contamination of the sample in the laboratory during the sample preparation.

This is a common problem when radiocarbon dating old samples, or during ^{36}Cl sample preparation in a laboratory in which neutron-activated samples were processed previously.

To finish this short overview of the basics of sample preparation we give a numerical example for ^{10}Be in ice. Table 10.3.3-1 shows that the calculated mean global production rate of ^{10}Be is $0.021 \text{ atoms cm}^{-2} \text{ s}^{-1}$ or $6.63 \times 10^5 \text{ cm}^{-2} \text{ y}^{-1}$. We use the rounded number of 1 million ^{10}Be atoms $\text{cm}^{-2} \text{ y}^{-1}$. These atoms are found in the mean global annual precipitation rate (rain and snow) of $\sim 100 \text{ cm y}^{-1}$. This corresponds to a concentration of $\sim 10^4$ ^{10}Be atoms per gram of ice or precipitation. Adding 1 mg of ^9Be corresponds to 6.69×10^{19} ^9Be atoms. If we round this number to 10^{20} ^9Be atoms the $^{10}\text{Be}/^9\text{Be}$ ratio in the sample is 10^{-16} , which is about 2 orders of magnitude below the detection limit of a present-day AMS system. The ratio can be increased easily by taking a larger sample and/or a smaller spike. The best strategy is to first choose a lower limit of the ratio which will provide a result with an acceptable statistical uncertainty within a reasonable measuring time. If we decide, for example, that the $^{10}\text{Be}/^9\text{Be}$ ratio should be 10^{-13} , this means that we need samples of 1,000 g when using a spike of 1 mg or samples of 100 g when using a spike of 0.1 mg, or anything in between. One may be tempted to choose the solution with the small spike. However, it is then necessary to take into account that this results in a smaller sample size, a possibly lower Be current from the ion source, and a faster burn out of the sample.

15.4 Decay Versus Atom Counting

The successful development of the accelerator mass spectrometry technique does not mean that the decay counting technique is outdated. As discussed in the previous section, AMS is the best method for long-lived radionuclides. In the case of short-lived cosmogenic radionuclides, it is generally still easier and cheaper to use decay counting. So, where do we draw the line between decay counting and accelerator mass spectrometry? In practice, there is no sharp line and the decision may be influenced by different criteria and the weights they are given. In both methodologies, we need to determine the number of atoms of the nuclide of interest in a given sample. Both methodologies yield results that follow essentially Poisson distributions and, in general, we will need to detect several hundred to several thousand atoms of the nuclide of interest to provide adequate statistical accuracy.

In the case of low-level counting, Fig. 15.4-1 displays the percentage of the radionuclide atoms in the sample that decay during different measuring times as a function of their half-lives. Assume that the design of the experiment indicates that there will be about 10^8 atoms of a cosmogenic nuclide in the sample. To count about 10^4 decays (to get adequate statistics), we will need to record the pulses from 10^4 atoms, that is 0.01% of the atoms in the sample. Figure 15.4-1 shows how long it

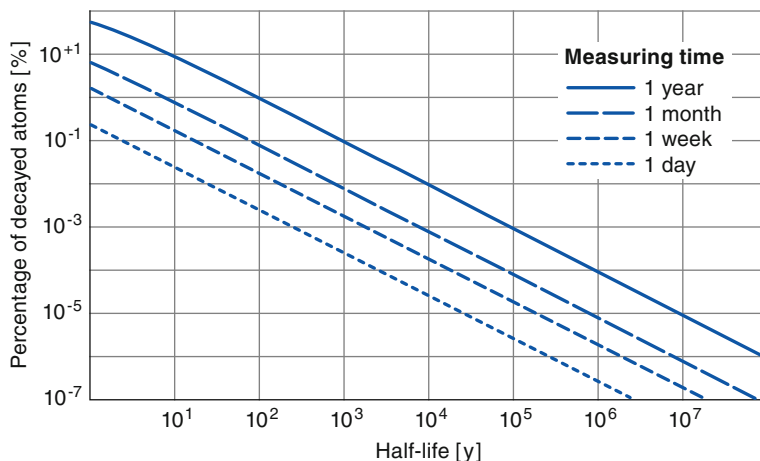


Fig. 15.4-1 Efficiency (percentage of the target atoms that decay) of decay counting versus the half-life and the measuring time. The longer the half-life and the shorter the measuring time, the smaller is the percentage of the atoms that decay during the measurement and the lower the efficiency of decay counting

will take to achieve that result. For a nuclide of half-life = 30 years or less, adequate statistics will be obtained by counting for a day or less. For a half-life of 1,000 years, it would take 1 month. An experiment yielding samples with ten times the number of radionuclides would require 0.001% of the nuclides to decay which would occur in 1 month for a half-life of 10,000 years.

In the case of AMS, it is also impossible to detect all the atoms in a sample. The number of atoms reaching the final detector is determined by the efficiency of the AMS, which is given by the yield of negative ions in the ion source times the fraction that then successfully travels through the accelerator mass spectrometry system into the detector, times the efficiency of the detection system. The losses are due to the different physical processes involved in the AMS. The first and most important source of loss occurs in the generation of negative ions in the ion source. For obvious reasons, it is much easier to remove electrons to produce positive ions than to attach an electron to a neutral atom. The yield of negative ions depends strongly on the element, the design of the ion source, the caesium temperature that mainly defines the intensity of the primary beam and several other parameters. Typical yields are in the order of a few percent (Cl, C) to almost one permil (Be, Al). The main factor determining the efficiency of transmission through the AMS is the stripping yield of the selected charge state at the positive terminal. The yield of the chosen charge state depends on the energy and the stripper medium (foil or gas) and is typically 50%. Overall, the efficiency of an AMS is typically 10^{-4} – 10^{-2} , strongly depending on the radionuclide and the AMS system, measurement technique and so on. For the assumed sample containing 10^8 atoms of the nuclides of interest, between 10^4 and 10^6 will reach the detector, which is adequate for the desired statistical accuracy. It is important to note that this result is independent of

the half-life of the nuclide, and essentially the same accuracy is attained for very short and very long half-lives, as illustrated in the following examples. The measuring time for a sample depends on the beam intensity leaving the ion source, the transmission, the isotopic ratio, and the desired statistical precision, and ranges typically from a few minutes to less than an hour.

The plot in Fig. 15.4-1 allows us to ask the question: “what are the circumstances under which low-level decay counting is competitive with an AMS system with an overall efficiency of 10^{-4} , say?” The figure shows that the 10^{-4} ($10^{-2}\%$) line corresponds to a counting time of 1 week for a radionuclide with $T_{1/2} = 200$ years, or ~ 1 month for $T_{1/2} = 1,000$ years. This indicates that the line between decay counting and accelerator mass spectrometry lies somewhere between half-lives of several hundred to a few thousand years.

To illustrate the different detection techniques we now give three examples of practical importance, where we are comparing low-level decay counting with AMS for which we assume an efficiency of 10^{-4} , this being the lower limit encountered in practice.

^{10}Be ($T_{1/2} = 1.387 \times 10^6$ years): Fig. 15.4-1 shows that the measuring time for the 10^{-4} line would be about 100 years. This shows why measurements of ^{10}Be were not possible in ice samples prior to the development of the AMS technique.

^{14}C ($T_{1/2} = 5,730$ years): The figure shows that for a sample containing 10^8 radioactive atoms, it would take almost a year to count 10^{-4} (0.01%) of them. However, if there were a factor of 30 more target atoms in the sample, adequate statistics would be reached by counting 3×10^{-6} of them, which would take a week. This example illustrates that ^{14}C is on the dividing line; and both low-level decay counting and AMS have been used to great effect in the past. Since the overall AMS efficiency for radiocarbon is closer to 10^{-2} than 10^{-4} , the vast majority of ^{14}C analyses are done today by AMS.

^7Be ($T_{1/2} = 53.2$ days): The graph shows that the 10^{-4} efficiency of the AMS would be attained in decay counting in a period of about 10 min. However, since only 10.5% of the decays of ^7Be produce a gamma ray, the measuring time is in reality about 100 min. This, and the simple and fast sample preparation required, explains why low-level counting is invariably used for this cosmogenic isotope.

15.5 Other Techniques, Optical Methods

The “weakest” points of the present tandem accelerator mass spectrometry systems are the fact that the ion source has to provide negative ions, which is more difficult than the production of positive ions. In fact, it is impossible for noble gases such as ^{39}Ar and ^{81}Kr , which have exciting applications in oceanography (^{39}Ar) and in ground water dating (^{81}Kr). Another “weakness” of the present ion sources is that they are only able to discriminate between certain elements, such as ^{14}N and ^{14}C or ^{26}Mg and ^{26}Al , because some elements do not form stable negative ions (N, Mg). Whenever an element present in the sample forms negative ions, it will be fed into

the AMS system and has to be suppressed later on by magnetic or electric filters. The problem that some elements do not form negative ions can be solved by using an AMS system which employs a cyclotron instead of a tandem.

An alternative approach to AMS is to achieve better discrimination against isobaric and molecular background in the ion source itself, which reduces the effort necessary to filter out the isotope of interest. In that case an accelerator is not needed anymore (Lehmann et al. 1986). Resonance ionization spectroscopy (RIS) is one such technique. In this technique, a gas containing the target atoms is selectively ionized in several steps using laser beams with quantum energies which match the energy gap between two energy levels in the target atoms. This discriminates against all other elements because these gaps are unique for the element of interest. The ions are then accelerated in a static electric field and separated using conventional mass spectrometry. To the best of our knowledge, this technique is not used frequently yet in the measurement of cosmogenic radionuclides.

15.5.1 *Final Remarks*

We began this chapter on detection by pointing out the importance of the analytical methods in the field of cosmogenic radionuclides. In the last 3 decades AMS became the standard method for nuclides with half-lives longer than ~100 y (^{32}Si) to ~1,000 y (^{14}C) with a few exceptions (noble gases) (Kutschera 2005). Today, there is a clear trend to reduce the size of tandem accelerator mass spectrometry systems (Suter 2004). This is possible due to improvements in the suppression of the isobaric and molecular backgrounds and the application of thinner windows and better detection systems. In the case of ^{14}C , the terminal voltage can be reduced from about 5 MV to about 0.2 MV which significantly reduces the size and complexity of such an AMS system. Further improvements in the yield of the negative ions in the ion source are to be expected in the future.

For further discussion of the AMS technique, and its applications, the reader is directed to the following publications:

Finkel RC, Suter M (1993) AMS in the earth sciences: technique and applications, in advances in analytical geochemistry, edited, pp. 1–114.

Tuniz C, Bird R, Fink D, Herzog G (1998) Accelerator mass spectrometry: ultra-sensitive analysis for global science. CRC Press, USA (1998) p. 371.

References

- Alvarez LW, Cornog R (1939) Helium and hydrogen of mass 3. Phys Rev 56:613
Anbar M (1979) The limitations of mass spectrometric radiocarbon dating using CN^- ions, paper presented at First Int. AMS Conference

- Bennett CL, Beukens RP, Clover MR, Gove HE, Liebert RB, Litherland AE, Purser KH, Sondheim WE (1977) Radiocarbon dating using electrostatic accelerators: negative ions provide the key. *Science* 198:508–510
- Kutschera W (2005) Progress in isotope analysis at ultra-trace level by AMS. *Int J Mass Spectrom* 242(2–3):145–160
- Lehmann BE, Loosli HH, Oeschger H, Rauber D, Hurst GS, Allman SL, Chen CH, Kramer SD, Thonnard N, Willis RD (1986) Laser resonance ionization mass spectrometry for Krypton-81 analysis. *Radiocarbon* 28(2):223–228
- Muller RA (1977) Radioisotope dating with a cyclotron. *Science* 196:489–494
- Nelson DE, Korteling RG, Stott WR (1977) Carbon-14: direct detection at natural concentrations. *Science* 198:507–508
- Purser KH, Liebert RB, Litherland AE, Beukens RP, Gove HE, Bennett CL, Clover MR, Sondheim WE (1977) An attempt to detect stable n^- ions from a sputter ion source and some implications of the results for the design of tandems for ultra-sensitive Carbon analysis. *Revue de Physique Appliquée* 12:1487–1492
- Suter M (2004) 25 years of AMS – a review of recent developments. *Nucl Instrum Methods Phys Res B* 223:139–148

Part IV

Applications

Chapter 16

Introduction to Applications

In the two previous Parts we have seen how cosmic ray particles (protons, alpha particles, and heavier nuclei) are accelerated almost to the speed of light when a massive star nearing the end of its life explodes as a supernova. The cosmic ray particles then travel in a “random walk” through their home galaxy for millions of years before interacting with matter or escaping into inter-galactic space. For those produced in our own galaxy the average residence time is six million years. Those of them that approach planet Earth must overcome two barriers before they can enter our atmosphere.

The first barrier is formed by the solar wind which continuously streams away from the Sun and forms a huge bubble, the heliosphere, which extends to approximately 100 times the distance between Sun and Earth. Not all the cosmic ray particles manage to penetrate the heliosphere. Scattering on “frozen-in” magnetic fields, convection and diffusion processes and deceleration cause a severe reduction in the intensity of the radiation, the magnitude of the decrease depending strongly on the magnetic activity of the Sun. Those cosmic rays that successfully make their way through the heliosphere encounter a second barrier, the geomagnetic field. Only those cosmic rays with a relatively high momentum to charge ratio can penetrate this magnetic shield and reach the atmosphere.

Then Part 3 has outlined the processes that occur once cosmic radiation enters the atmosphere of Earth. We have seen that a cascade of nuclear reactions sets in leading to the production of a variety of cosmogenic radionuclides. Most of them are very short-lived and quickly decay into stable atoms which cannot be distinguished from those of the same type already present in the environment. The longer-lived cosmogenic radionuclides, however, take part in all the various processes listed in Fig. 3.1. Since they carry a “colour” and a “clock” they offer the opportunity to study many transport processes and provide information on the time of occurrence of events in the past when stored in natural archives.

Having described these theoretical concepts of what cosmogenic radionuclides are in the previous Parts of this book, it is now time to see what they are good for, and to discuss their potential in practical applications. Instead of trying to give a complete overview of what has been done so far, we have selected a number of typical examples from the main subsystems of the environmental system (Fig. 3.1). It is our hope that this will provide the reader with a deeper understanding of the subject and provide motivation to find new applications in other fields of research.

Chapter 17

Solar Physics

17.1 Introduction

Cosmogenic ^{14}C was first measured in the 1940s using the “low level decay counting” techniques outlined in Sect. 15.2, and was soon being used for “carbon dating” (Chap. 23). This was before Scott Forbush discovered the 11-year variation in the galactic cosmic radiation (Sect. 7.2), and it was assumed that the production rate of the cosmogenic radionuclides in the atmosphere was independent of time. In 1957 however, following Forbush’s discovery, Bernard Peters and Devendra Lal (1962) predicted that the ^{14}C production rate would vary with time. In the same year Walter Elsasser, a “solid earth” geophysicist, and two cosmic ray physicists, Ed Ney and John Winckler, predicted that the ^{14}C concentration would be affected by the changing strength of the geomagnetic field. Soon after, de Vries obtained experimental evidence that the ^{14}C production rate had been higher in ~1700 AD than in the nineteenth century. Several years later Hans Suess demonstrated that there had been several variations of the ^{14}C production rate over time, and these became known as the “Suess wiggles” (Suess 1970).

As described in Sects. 5.7 and 5.8, we now have a good understanding of both the solar and geomagnetic effects. We understand the manner in which the heliospheric and geomagnetic fields influence the intensity and spectrum of the cosmic radiation entering the Earth’s atmosphere, and how they have controlled the rate of production of the cosmogenic radionuclides in the past (Chap. 10). As a consequence, we can use the paleo-cosmic ray “Rosetta Stone” (Sect. 5.2) and the cosmogenic radionuclide data to investigate the physics of the Sun, the heliosphere, and the geomagnetic field in the past.

While mankind’s detailed observations of solar and geomagnetic phenomena commenced only two to four centuries ago, the cosmogenic archives allow us to extend our knowledge a further 30,000–100,000 years into the past. This provides a much better understanding of the nature of the long-term changes, which then allows us to estimate the speed and magnitude of changes that may occur in the future. Some of our modern technology, such as satellites, communication, and

power systems, are adversely affected by some aspects of solar activity. A better understanding of the past allows us to compute risk and the financial consequences of changes that may occur in the future.

In this chapter, we therefore “turn the problem around” and outline how the cosmogenic radionuclide data have been used to investigate the changes in the Sun over the past 10,000 years. We then briefly outline the practical applications and implications of that knowledge. Chapter 21 will look at the long-term changes in the geomagnetic field in a similar manner.

17.2 Solar Periodicities and the “Grand Minima” in the Cosmogenic Radionuclide Record

17.2.1 Solar Periodicities: Time Domain Studies

Cosmogenic radionuclide and cosmic ray measurements are always made in the form of a time series, ranging from the earliest date to the latest or vice versa. Such data are said to be measurements in the “time domain”. Figures 5.5-2 and 6.2-3 and all the examples in Chap. 7 are “time domain measurements”, although it is not necessary to say this since it is obvious from their presentation. It is possible to convert time domain data to the “frequency domain”, which is ideally suited for recognizing the presence of hitherto unrecognized periodicities. Fourier and power spectrum analyses are two such methods that will be briefly mentioned in the next paragraph, and in more detail in Box 17.2.1.1 and Sect. 17.2.2

As outlined in the introduction, the ^{14}C data provided the first direct evidence that there were temporal changes in the rate of production of the cosmogenic radionuclides. As more data were accumulated from the analysis of tree rings, it became apparent that there was a ~208-year periodicity in the ^{14}C record (the “Suess” or “De Vries” cycle). This initial result was verified using longer and more accurate ^{14}C records, and through Fourier and power spectrum analysis. The Gleissberg periodicity at ~80–90 years (previously identified in solar and geophysical data) and a ~2,300-year periodicity were also detected in the ^{14}C data using Fourier and power spectrum analysis (Damon and Jirikowic 1992; Stuiver et al. 1991). The presence of these periodicities was later verified using ^{10}Be data (see Box 17.2.1.1).

17.2.1.1 BOX Period, Frequency, Harmonics, and Frequency Domain Analysis

A “periodic signal” is one that repeats itself after an interval of time, T , called the “period” of the signal. The simplest periodic signal is the sine function,

$$f(t) = A \sin(2\pi t/T) = A \sin(2\pi f t) \quad (\text{B17.2.1.1-1})$$

where A is the amplitude of the variation, t is time, π is the constant 3.1416, T is the period, and the frequency of the variation is defined as $f = 1/T$. In nature, we often see “quasi-periodic signals”, where the amplitude A may vary from one cycle to the next, as may the period and frequency.

Before proceeding, we note that there can be confusion regarding the meaning of the word “amplitude”. To the mathematician, there is no confusion – it is A as defined in (B17.2.1.1-1). That is, $f(t)$ varies from plus A to minus A . However, in common usage the amplitude of a variation often means the total excursion from the lowest value of a quasi-repetitive signal to the highest value; this is also called the “peak-to-peak” amplitude. That is, the “peak-to-peak” amplitude of $f(t)$ in (B17.2.1.1-1) is $2A$. For long-term cosmic ray records such as those shown in Figs. 7.2-1 and 7.3-1, amplitude usually refers to the “peak-to-peak” amplitude.

The 11-year cycles in the cosmic ray intensity and sunspot number in Fig. 6.2-3 are good examples of “quasi-periodic” signals. They both have a series of maxima and minima at a separation of approximately 11 years. The figure shows that the amplitudes vary between cycles. Close examination shows that the period varies by up to a year or more from cycle to cycle as well. The mean period of these signals is ~ 11 years and the frequency $= 1/T = 0.0909$ cycles/year. The Suess periodicity has a period of 208 years and a frequency of 0.0048 cycles/year.

For a signal at the “fundamental” frequency f , ($A \sin(2\pi ft)$), the signal at twice the frequency, ($A \sin(2\pi(2f)t)$), is called the “second harmonic”. The signal at a frequency of $3f$ is the third harmonic, and so on. As is clear from the definition ($f = 1/T$), the period of the second harmonic is half that of the fundamental. Thus, the period of the second harmonic of the 208-year Suess periodicity is 104 years.

The mathematical process, the “Fourier Transform”, allows a time series $X(t)$, such as in Figs. 6.2-3, 7.3-3 and 17.2.1-1 to be expressed as its “frequency domain” equivalent. In the frequency domain, the observed data are represented by a series of periodic waveforms, which, when added together, yield the observed data. Thus,

$$X(t) = \sum_{n=1}^{n=\infty} A_n \sin\left(\frac{2\pi t}{T_n}\right) \quad (\text{B17.2.1.1-1})$$

This is a very powerful technique for detecting weak periodic signals that are buried in other signals and random noise, as discussed in the main text. There are other frequency domain techniques such as the discrete Fourier transform (DFT), the power spectrum, and the wavelet transform that provide a similar function.

Figure 17.2.1-1 displays the tree ring ^{14}C observations covering the past 12,000 years. Before proceeding, we note that the ^{14}C data in the literature will be found quoted in two quite different ways:

1. The first is where the observations are quoted relative to a standard value, yielding the quantity $\Delta^{14}\text{C}$ defined in Sect. 23.2.1. Figure 17.2.1-1 is such a case. As discussed in Sect. 13.5.3.2, the carbon cycle strongly attenuates the periodicities in the ^{14}C production rate, e.g. by a factor of 100 for the 11-year cycle and a factor of 10 for the 2,300 years (Hallstatt cycle). As a consequence, the $\Delta^{14}\text{C}$ record severely under-represents the short period variations in the cosmic ray intensity in the past.
2. The second manner of presentation is fundamentally different in concept. Here a mathematical process (see Sect. 13.5.3.4) is used to compute the production of ^{14}C (at the time of interest in the past) which, after attenuation by the carbon cycle, has resulted in the observed value of $\Delta^{14}\text{C}$. Figure 17.2.1-2 is an example. This process automatically removes the attenuation effects seen in the $\Delta^{14}\text{C}$ data, and the amplitudes of the several periodicities are presented in an unbiased manner. The production estimate using this method is sometimes written as $Q^{14}\text{C}$.

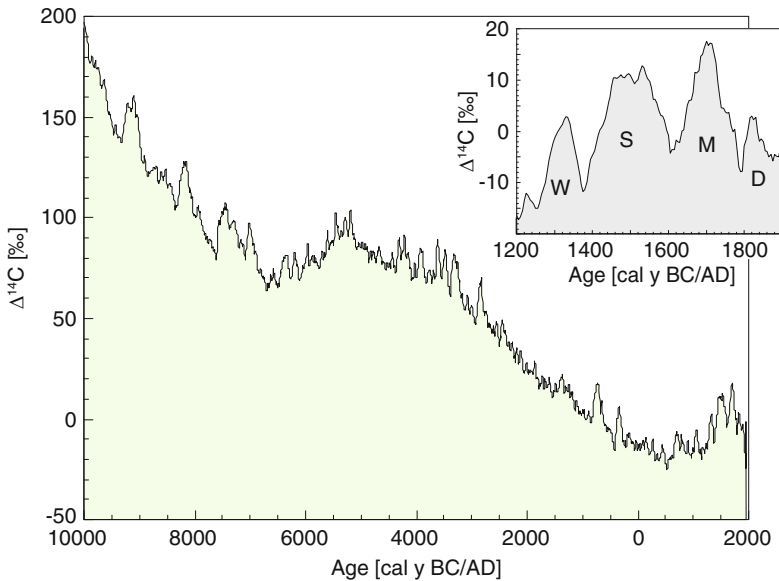


Fig. 17.2.1-1 The $\Delta^{14}\text{C}$ data corresponding to the interval 10000 BC – 2000 AD (McCracken et al. 2005). Note that the century-scale variations in the *inset* are similar to those in Fig. 7.3-1, corresponding to the Wolf (W), Spörer (S), and Maunder (M) and Dalton (D) Minima. These and the many other century-scale variations are superimposed on a changing baseline due to the long-term changes in the cosmic ray intensity (Sect. 17.3.2) and the geomagnetic field (Sect. 21.2). The quantity $\Delta^{14}\text{C}$ is defined in Sect. 23.2.1; note that these data are measured in permil (‰ = one part in a thousand)

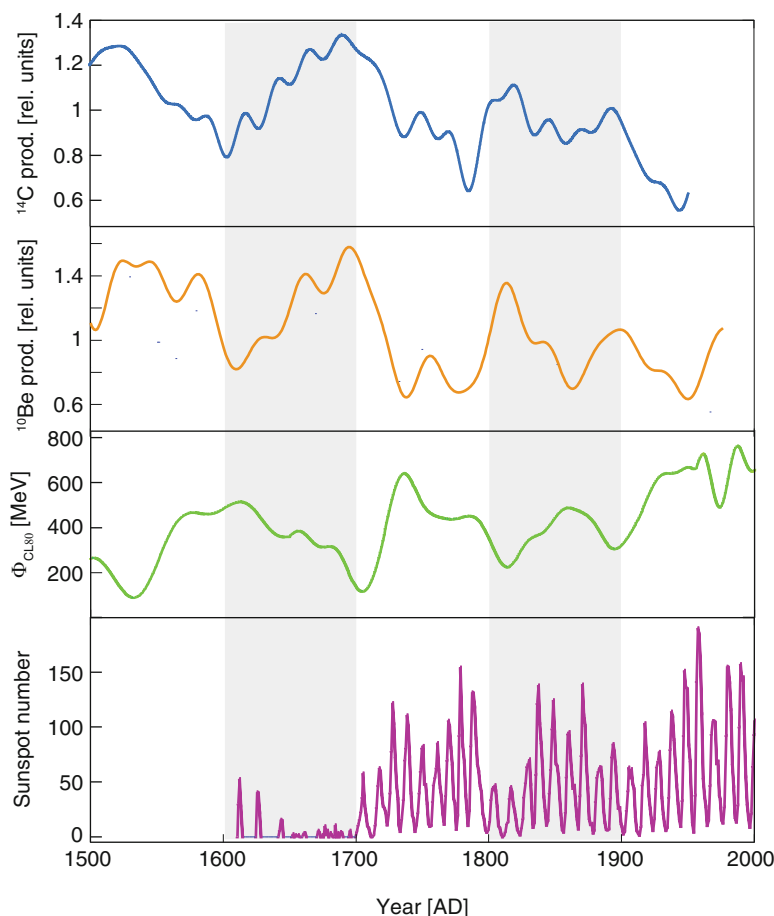


Fig. 17.2.1-2 Illustrating the inverse relationship between the cosmogenic radionuclide data and solar activity. The ^{14}C and ^{10}Be production rates are high during the periods when the solar cycles are weak (i.e. low maximum sunspot numbers), such as during the Maunder and Dalton Minima as shown. ^{14}C production derived from annual $\Delta^{14}\text{C}$ (Stuiver and Braziunas 1993), ^{10}Be from South pole record (Bard et al. 1997), Φ_{CL80} from (Steinhilber et al. 2008). ^{10}Be and ^{14}C production rates are low-pass filtered with 22 y, Φ_{CL80} with 40 y

In Fig 17.2.1-1 each data point is averaged over 10 years and the 11-year solar cycle is averaged out, first by the storage effects in the carbon cycle (Sect. 13.5.3), and secondly in the sampling process (Sect. 14.6.1). There are two striking features of this record: (1) the slow decline (with temporary reversals starting ~8000 BP and ~2000 BP), and (2) the occasional high frequency oscillations (periods ~200 y) superimposed thereupon. The first of these is attributed to the slowly changing strength of the geomagnetic field and is discussed in Sect. 21.2. The higher frequency variations are discussed in this and the next sections.

The solar physicist John Eddy (1976) recognized that the high frequency oscillations in the $\Delta^{14}\text{C}$ record were out of phase with the long-term variations in the amplitude of the sunspot cycles from the commencement of modern solar observations in 1610. Figure 17.2.1-2 displays this anti-correlation for both $\Delta^{14}\text{C}$ and ^{10}Be (see also Fig. 7.3-1). As discussed in Sect. 5.5, there were virtually no sunspots on the Sun throughout the “Maunder Minimum”, 1645–1715, and the figure shows that the ^{14}C and ^{10}Be concentrations were high. The maximum sunspot numbers attained during the low amplitude sunspot cycles in the vicinity of 1810 (“the Dalton Minimum”) and circa 1900 were also low, and the average ^{14}C and ^{10}Be were high again. The occurrence of several solar cycles with low peak values of solar activity (low maximum sunspot number) is called a “Grand Minimum”, so the hypothesis was put by Eddy that high values of ^{14}C corresponded to Grand Minima in solar activity. Figure 17.2.1-2 (which includes data that was not available to Eddy) shows that this is a semi-quantitative relationship: the lower the solar activity, the higher the production of the cosmogenic radionuclides. Thus, the Maunder Minimum was deep (very low sunspot numbers), the Dalton Minimum less pronounced (peak sunspot numbers ~ 30), and the Gleissberg Minimum of 1900 was the weakest of the three (peak sunspot number ~ 60). The ^{10}Be concentration and ^{14}C production proceed in the reverse order. They are highest for the Maunder Minimum, lower for the Dalton, and least for the Gleissberg Minimum. On the basis of the 400 years of concurrent solar and ^{14}C data available at the time, Eddy concluded that there was an empirical anti-correlation between the ^{14}C data and the level of solar activity.

Based upon this anti-correlation, Eddy examined peaks in the ^{14}C record in the vicinity of ~ 1050 AD, ~ 1250 AD, and 1420–1540 AD. He concluded that the Chinese sunspot records supported the hypothesis that these earlier peaks in ^{14}C corresponded to periods of low solar activity. Furthermore, the Chinese records showed that sunspots were quite common during the interval 1100–1250 AD, and again in 1325–1400 AD, corresponding to the minima between the ^{14}C peaks. Steadily improving ^{14}C data and the ^{10}Be data obtained after 1980 have completely verified the validity of the anti-correlation hypothesis.

This anticorrelation is explicable in terms of the solar modulation of the galactic cosmic radiation described in Sect. 5.7. As described there, the cosmic ray intensity and energy spectrum are under solar control – the more active the Sun, the stronger the “modulation” of the cosmic radiation – leading to reduced production rates of the cosmogenic radionuclides. These changes can be quantified in terms of the modulation function $\Phi_{\text{CL}80}$ (see Sects. 5.7.3 and 5.7.4), as shown in Fig. 17.2.1-2. Thus, the lowest value of $\Phi_{\text{CL}80}$ corresponds to the very low solar activity during the Maunder Minimum; it is somewhat higher for the Dalton and Gleissberg Minimum of 1900, and is highest during periods of high solar activity such as near 1750, 1850, and from 1950 onwards.

The geomagnetic field was relatively stable during the interval shown in Fig. 17.2.1-2 and changes in the production rates of the ^{14}C and ^{10}Be were nearly all of solar origin. In general, however, the geomagnetic modulation must be removed first to leave an unambiguous solar signal. The geomagnetic effects are described further in Chap. 21.

17.2.2 *Solar Periodicities: Frequency Domain Studies*

As discussed earlier, the first evidence for repetitious variations in the ^{14}C data was obtained by detection of a period of ~ 208 years in the data. Several mathematical techniques are commonly used to detect periodicities: Fourier analysis, digital Fourier transforms (DFT), power spectra, and wavelet analysis are four such. These studies – said to be in the “frequency domain” – are to be contrasted with the studies in Sect. 17.2.1 that are said to be in the “time domain” (see Box 17.2.1.1). We stress, however, that the frequency and time domain descriptions are equally accurate ways to describe the same data and that there is exact mathematical equivalence between the two descriptions. In practice, analysis in the frequency domain is an ideal way to detect previously unsuspected periodicities in data, as occurred in the discovery of the 208-year periodicity, and later the $\sim 2,300$ -year periodicity. On the other hand, it is not easy to use frequency domain studies to determine the associations between different data, e.g. the identification of the close association between sunspot numbers, and the cosmogenic radionuclide data discussed in Sect. 17.3.1, or to identify the sporadic nature of the episodes of grand minima (Sect. 17.3.2). Analysis in the “time domain” is the more appropriate way to perform such studies.

That is, frequency domain methods are “the method of choice” in appropriate applications and this section provides a summary of the results obtained with these methods. We start with a frequency domain analysis of 9,300 years of the time domain ^{10}Be data shown in Fig. 17.3.2-1, followed by analyses of the 12,000 years of $\Delta^{14}\text{C}$ data shown (in the time domain) in Fig. 17.2.1-1, and the 9,300 years of $Q^{14}\text{C}$ data shown in Fig. 17.3.2-1. In this way, you can contrast the insight gained using the time and frequency (yet, mathematically, both equally valid) methods of analysis when applied to exactly the same data sets, and also examine the agreement between these two cosmogenic radionuclides. As we have seen in Chap. 13, the atmospheric transport processes suffered by ^{10}Be and ^{14}C are quite different, and this comparison gives us confidence that the periodicities in the data are due to variations in the cosmic radiation incident on Earth (i.e. they are “production” effects) and that they are not “statistical noise” or “system effects”.

The Fourier spectrum in Fig. 17.2.2-1 is a mathematical representation in which the original “time domain” data set in the second panel of Fig. 17.3.2-1 is described exactly, as a sum of sine waves at the frequencies given along the x -axis (see Box 17.2.1.1). The amplitude of the sine wave at any given frequency is given by the height of the vertical line at that frequency.

The first impression gained from Fig. 17.2.2-1 is that the mathematical technique appears to be saying that there are signals at hundreds of different frequencies in the original data. This is not correct; however, the mathematics has faithfully represented all the statistical noise in the data (as well as the experimental errors, etc.) as well as the signals that are there. The amplitude values for many of the frequencies in Fig. 17.2.2-1 are not statistically significant (see Box 17.2.2.1).

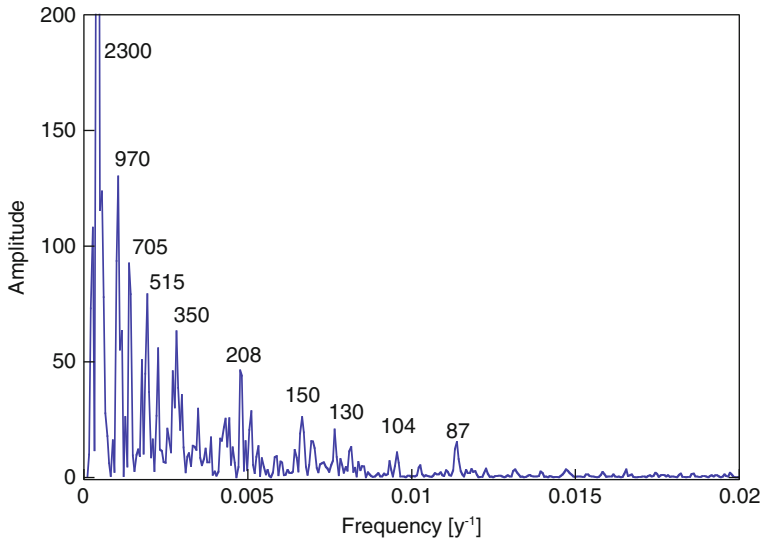


Fig. 17.2.2-1 Fourier spectra of the time domain ^{10}Be data shown in Fig. 17.3.2-1. The x-axis is the frequency corresponding to the periodicities in the spectrum, and the y-axis is the amplitude of the Fourier components used to describe the data. The periods of the major peaks are annotated in years. The presence of noise, harmonics, and heterodyne frequencies in this spectrum is discussed in the text

17.2.2.1 BOX Signals, Noise, Numerical Filters, and Statistics

The cosmogenic radionuclide data that we use contains various temporal variations due to external factors (solar, geomagnetic, and climate to name three), systematic errors (e.g. year-to-year variations in snow fall), and statistical noise introduced by the sampling and analytical techniques. Depending on our interests we may call some of these variations “signals” and all the other variations “noise”. For example, the “solar signal” of interest to the solar physicist will interfere with other applications of the cosmogenic radionuclide data; in such a case the solar signal represents “noise” that can obscure or corrupt the “signals” of interest to another scientist (e.g. in carbon dating). The measurement errors, etc. may be comparable in size to the signals of interest and it is invariably necessary to use statistical techniques to distinguish between the signal of interest and the noise.

Further, it is then necessary to use statistical tests to determine if the result so obtained is “statistically significant”. That is, to determine whether the supposed result is only due to random fluctuations in the data, or due to the other signals in the data (e.g. solar, geomagnetic, climate, etc.), that have been allowed to masquerade as the signal of interest in the mathematical process used (e.g. “aliasing”, heterodyning, etc.).

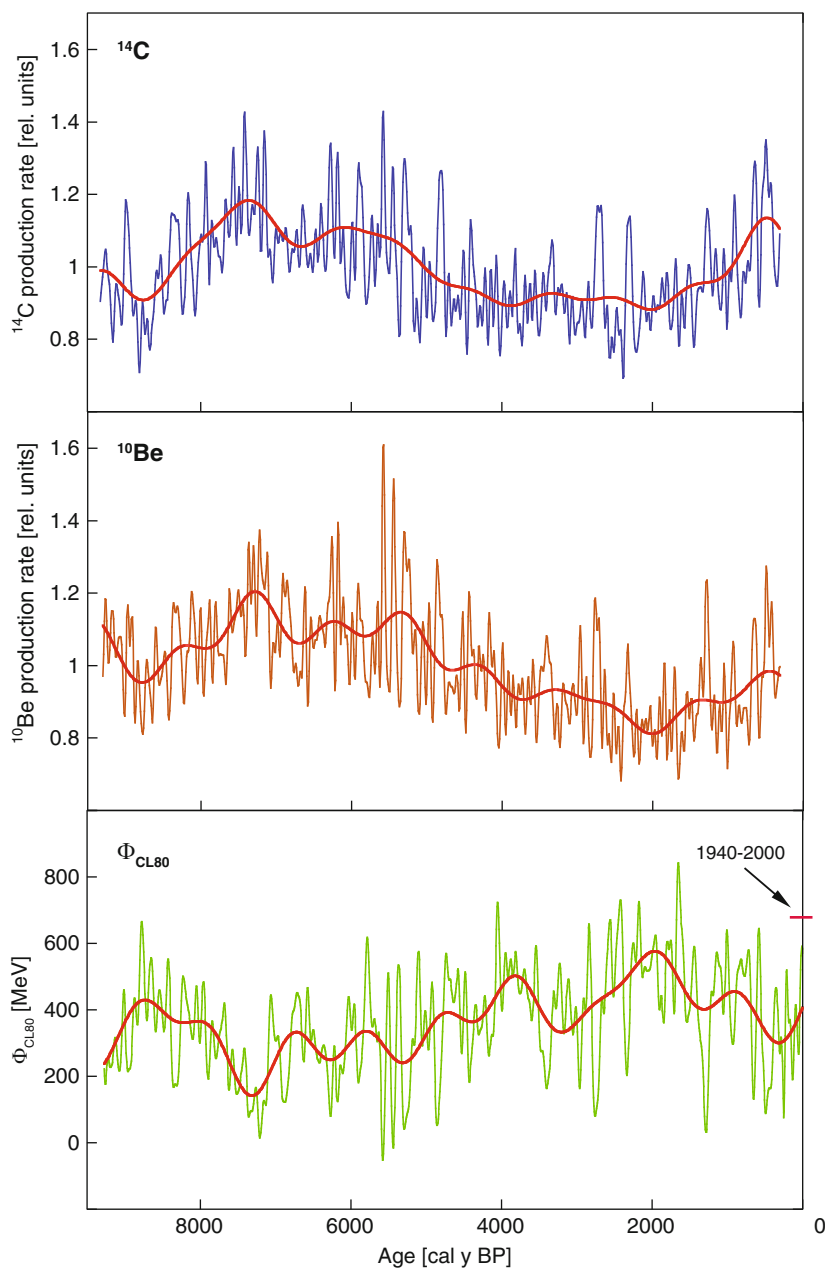


Fig. 17.3.2-1 The cosmogenic radionuclide data for the past 9,300 year and the deduced cosmic ray modulation function. All the data have been filtered to remove the 11- and 22-year variations. Panels (a) and (b), the ^{14}C and ^{10}Be production rates normalized to a mean production rate of 1; panel (c), the modulation function derived from a principal components analysis of the data in panels (a) and (b). Note that the time axis is “time before present”. ^{10}Be from GRIP, low-pass

It is well known that we can improve the accuracy of any measurement by averaging a number of independent measurements of that quantity. In the same way, we can reduce the “noise” in a time series, $x_1, x_2, x_3, \dots, x_n$, by averaging a number of successive values. This is referred to as “applying a numerical filter to the data”. The “binomial filter” is frequently used for this purpose where different “weights”, $c_1, c_2, c_3, \dots, c_n$, are used to obtain the “filtered value” corresponding to x_3 as

$$X_3 = (c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5)/(c_1 + c_2 + c_3 + c_4 + c_5)$$

where, for this particular case, the weights are (1,4,6,4,1). Other binomial filters in frequent use in cosmogenic radionuclide data analysis are those specified by the coefficients (1,2,1) and (1,3,3,1). Note that the numerical filter removes higher frequency components from the data stream. If the data were annual ^{10}Be data, the (1,4,6,4,1) filter will largely eliminate any variations with a period up to 3 years, while leaving the 11-year variation essentially unaffected.

“Signal Processing” is the modern term for the process of identifying and measuring weak signals (i.e. what we are interested in) in the presence of random errors and other signals (noise). This box merely stresses that there are many traps that await the unsuspecting scientist in this process. It is well to remember the two sayings: (1) about statistical inference – “there are lies, damned lies, and then there are statistics” – reminding us that statistical black magic has the potential to make us believe what we should not; and (2) about signal processing – “garbage in, garbage out”.

It is also to be noted that there has been a “credibility” problem with frequency domain results. The sight of hundreds of frequencies such as in Fig. 17.2.2-1, and to be told that most of them “mean nothing”, often has the effect of making objective scientists doubt that the frequency domain methods tell them anything at all. To convince them, it is vital that sound statistical tests be applied that yield the “confidence limits” that the peak of interest in the power spectrum is “real”, and not just a response to random noise. Without those statistical tests, it is not worth your time to produce the frequency domain results in the first place.

The user of cosmogenic radionuclide data must therefore consult appropriate specialist texts and experts if they are to use the mathematical and statistical tools in an effective manner. It is also important to do so to be able to judge whether results published by other scientists are reliable or not.

Fig. 17.3.2-1 (continued) filtered with 60 y, ^{14}C from INTCAL04 with carbon cycle model, low-pass 60 y, Φ_{CL80} from Vonmoos and Steinhilber, low pass 60 years (Muscheler et al. 2004; Reimer et al. 2004; Steinhilber et al. 2008; Vonmoos et al. 2006). Red curves, 1,000-years low-pass filter. Note that the time scale ends at 0 BP (1950 AD). The mean level of solar activity between 1940 and 2000 AD is indicated in the Φ_{CL80} panel (670 MeV)

The labelled peaks, and some of the higher intermediate ones in Fig. 17.2.2-1, are statistically significant. The 208-year peak is the Suess periodicity (also called the de Vries periodicity) discussed previously. The 87-year peak corresponds to the Gleissberg periodicity known in sunspot and geophysical records (Sect. 5.5), the 970-year periodicity is dedicated to Jack Eddy (1976), and the 2,300-year peak has been named the “Hallstatt” periodicity by Damon and Sonett (1991).

As discussed in Sect. 13.5, the atmospheric transport mechanisms of ^{10}Be and ^{14}C are quite different, and consequently the atmospheric and meteorological system errors in these two data series are expected to be uncorrelated. Thus, the periodicities due to changes in the galactic cosmic ray intensity will appear in both records, while periodicities that originate in atmospheric transport effects (and ocean effects in the case of ^{14}C) will be in one but not the other. This provides a powerful test of whether the periodicities observed in the cosmogenic radionuclide data are due to “production” or “system” effects. To apply this test, Table 17.2.2-1 compares the spectral peaks obtained by frequency domain analysis of the ^{10}Be , $\Delta^{14}\text{C}$, Q^{14}C data and the modulation function (Φ_{CL80}) records. It is clear that there is very close agreement between the periodicities in the totally independent ^{10}Be and ^{14}C data, confirming that they are production effects, and a consequence of periodic changes in the cosmic ray intensity in the past.

The time scales for ^{10}Be in ice are not as easily established as those for ^{14}C in tree rings, and the close agreement between the periods for ^{10}Be and ^{14}C provides verification of the absolute values of the computed periodicities. It is important to note that different frequency domain techniques, different “windowing” procedures (necessary to avoid spurious results due to “aliasing” and other effects), or the use of selected portions of the original record will all give slightly different values for these periodicities. In particular, note that spectral analysis software computes the spectral components for typically 512 or 1,024 equally spaced frequencies. As a consequence of the inverse relationship of frequency and period, this means that adjacent

Table 17.2.2-1 Spectral peaks derived from (1) the 9300 year GRIP ^{10}Be record in Fig. 17.2.2-1; (2) the $\Delta^{14}\text{C}$ record in Fig. 17.2.1-1; (3) a Q^{14}C record for the interval 1750–9300 BP; (4) a 9300-year computation of the modulation function based on the GRIP ^{10}Be data

Name	^{10}Be	$\Delta^{14}\text{C}$	Q^{14}C	Φ
Hallstatt	2,300	2,275	2,341	2,194
Eddy	970	984	985	982
	705	714	713	704
	515	512	512	497
	350	350	351	352
Suess	208	208	208	207
	150	150*	150	148
	130	130*	130	130
	104	105	105	105
Gleissberg	87	88	87	86

The periods are all given in years. The reader should note the clearly defined nature of the spectral peaks corresponding to each of these ten periods in Fig. 17.2.2-1. (Source: Beer and McCracken 2007). Periods marked * from Peristykh and Damon (2003)

frequencies correspond to periods that differ by ~ 120 years near the Hallstatt period, ~ 25 years near the Eddy period, and 0.8 years near the Suess period. Consideration of these factors indicates that the differences in period in the upper part of Table 17.2.2-1 are consistent with the characteristics of the software and the limited length of the data records compared to the periods under consideration.

As outlined in Box 17.2.1.1 it is common for naturally occurring periodicities to have second and higher harmonics. Furthermore, when two frequencies, f_1 and f_2 are added, and then passed through a non-linear physical or mathematical process, two additional “heterodyne” frequencies ($f_1 + f_2$) and ($f_1 - f_2$) are present in the output (sometimes called “sideband”, “cross-modulation”, or “beat” frequencies), together with the original input frequencies.

Remembering that frequency, $f = 1/T$, where T is the period, then the heterodyne periods corresponding to the non-linear combination of two periods of T_1 and T_2 are

$$T_{\text{sidebands}} = \frac{1}{f_1 \pm f_2} = \frac{1}{1/T_1 \pm 1/T_2} = \frac{T_1 \cdot T_2}{T_1 \pm T_2} \quad (17.2.2-1)$$

Looking at Table 17.2.2-1 and Fig. 17.2.2-1 then,

- The peaks with periods of ~ 104 years and ~ 45 years (*Peristykh and Damon 2003*) may be attributed to the second harmonics of the Suess (~ 208 years) and Gleissberg (~ 87 years) periodicities.
- The peaks at ~ 150 years (Fig. 17.2.2-1) and ~ 60 years (*Peristykh and Damon 2003*) may be attributed to the heterodyne frequencies between the Gleissberg and Suess frequencies (e.g. the upper and lower sidebands in this case have periods of $(208 \times 87)/(208 \pm 87) = 149.6$ and 61.3 years.

Table 17.2.2-1 and Fig. 17.2.2-1 have far-reaching implications for solar and heliospheric physics. As discussed in Sect. 5.7, the modulation of the GCR is due to time-dependent changes in the characteristics of the heliospheric magnetic field (HMF) and the solar wind. Therefore, the fact that the GCR intensity exhibits the periodicities in Table 17.2.2-1 implies that those periodicities are also present in the characteristics of the HMF and/or the solar wind. The solar magnetic fields are attributed to a “solar dynamo”, and the solar magnetic fields, in turn, give rise to the HMF. Therefore, the presence of the periodicities in Table 17.2.2-1 in the paleo-cosmic ray record implies that they are also present in the characteristics and outputs of the solar dynamo. The sharp nature of the spectral peaks in Fig. 17.2.2-1 and the good agreement evident in Table 17.2.2-1 suggest that the responsible physical processes are of constant frequency and are long lived.

Finally, we note that great care is necessary in the application and interpretation of frequency domain techniques. Failure to remove long-term trends in the data may result in the “discovery” of an infinite series of periodicities at T , $T/3$, $T/5$, $T/7$... in the data. Sideband frequencies must be examined with care. But, given that, the frequency domain methods provide estimates for periodicities in the data that are difficult to detect or measure in the time domain. Together, time and frequency domain studies provide insight that neither can provide easily by itself.

Figure 17.2.2-2 provides another and valuable insight into the manner in which the solar periodicities in paleo-cosmic ray record have varied over the past 10,000 years. Here, in concept, a numerical filter tuned to frequency f , say, has scanned the full 10,000 yrs of $Q^{14}\text{C}$ data in 1500 year blocks at a time, and determined the amplitude of the signals at that frequency in each block. This has then been repeated for all periodicities $T > 50$ yrs. Combining the results, the figure shows that the amplitudes of the periodicities in the ^{14}C production rate have varied with time. For example, as summarized in the caption, the Suess periodicity has exhibited periods of low and high amplitude, at intervals of ~ 900 and ~ 2300 y. Amplitude modulation generates heterodyne frequencies; and equation 17.2.2-1 applies. This equation shows that amplitude modulation of the Suess periodicity by the Hallstatt (2300 y) and Eddy (970 y) periodicities will yield sideband frequencies at 172, 191, 229, and 264 years. Examination of Fig. 17.2.2-1 shows a roughly symmetrical series of spectral peaks, centred on the Suess periodicity and corresponding to periodicities in the range 170–270 yr. Thus the periodicities in the vicinity of the Suess periodicity appear to be consistent with the amplitude modulation evident in Fig. 17.2.2-2. Our discussion in Section 17.3.2 will lead us to conclude that the amplitude modulation in Fig. 17.2.2-2 is associated with the “clusters” of Grand Solar Minima in the past, and that they are responsible for some of the sidebands evident in Fig. 17.2.2-1.

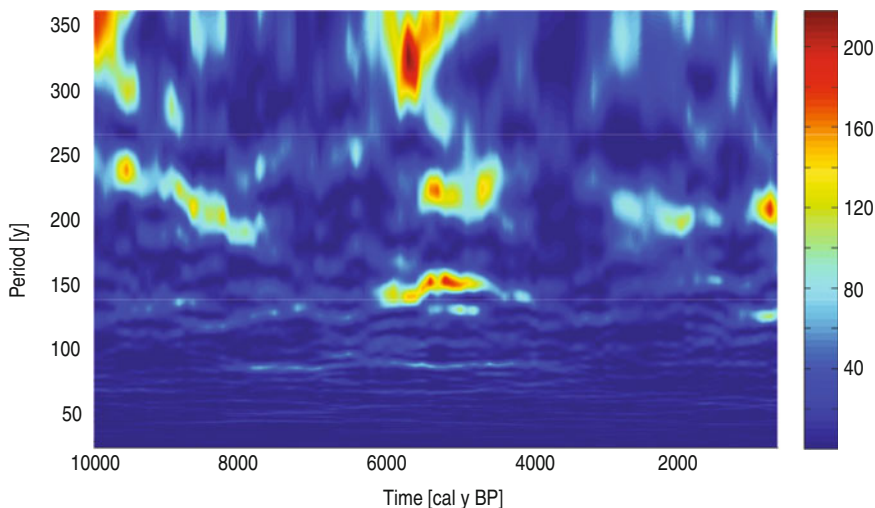


Fig. 17.2.2-2 A frequency domain plot of the variability of the rate of production of ^{14}C over the past 10,000 years. This shows that the amplitudes of the various periodicities have varied with time. For example, the Suess periodicity (208 years) was strongest in the intervals centred on ~ 500 , ~ 2000 , ~ 2800 , ~ 4600 , ~ 5500 , and ~ 7800 BP and weak in between. Note the strong variations in the amplitudes of the Gleissberg (87y), 150 y, and 350 y periodicities

17.3 Cosmic Ray and Solar Effects in the Past

17.3.1 *The Past Millennium*

In Sect. 6.6 we described how the modern instrumental measurements, and our theoretical knowledge, have allowed the cosmogenic radionuclide data to be inter-calibrated, and combined with the modern instrumental record. Using neutron monitor and ionization chamber data, together with cosmogenic radionuclide data, Figs. 6.2-3 and 6.6-2 have shown the manner in which the cosmic ray intensity has varied in a continuous manner over the past 600 years. Averaging over 22 years to eliminate the 11 and 22-year variations, Fig. 7.3-1 shows that the cosmic ray intensity and the solar activity have varied in a rather complex manner over the past 1,100 years. This provides a baseline for understanding the long-term solar variability discussed in the next section. It is yet another use of the cosmogenic “Rosetta Stone”, where the knowledge we have gained from the space age, the 400 years of high quality sunspot data, and several long ^{10}Be records provides us with the basis for understanding the more distant past.

In Sect. 17.2.1, it was shown that “Grand Minima” are accompanied by high production rates of cosmogenic radionuclides. Figure 7.3-1 shows that there were six peaks in the production rate of ^{10}Be between 850 and 2008 AD, indicating six Grand Minima of low solar activity during this period. The first three are called the Oort (~1050), Wolf (1325–1375), and Spörer (1420–1540) Minima and then follow the Maunder, Dalton, and the 1900 Gleissberg Minima which are also evident in the sunspot record (Sect. 5.5). The Spörer Minimum appears to have been the most profound – the cosmogenic radionuclide data indicate that it lasted for about a century, and that the cosmic ray modulation function (Φ_{CL80}) was very low, indicating a very low level of solar activity.

By way of contrast, Fig. 7.3-1 shows that the ^{10}Be concentration was very low for 2 centuries: 1075–1275, a period frequently called the “Medieval Maximum” (i.e. of solar activity). The right-hand vertical scale in Fig. 7.3-1 shows that the modulation function was high throughout this interval, the 22-year averages being in the range 500–700 MeV. Note that the 22-year average modulation function has been in this same range for the past 50 years (Fig. 7.3-1), indicating that the solar activity in the latter half of the twentieth century has been comparable to that in the “Medieval Maximum”. The sunspot record in Fig. 7.3-1 shows that the peak solar activity has been steadily increasing since the Maunder Minimum, and that the “space age” (1960–present) corresponds to the highest levels of activity since sunspot observations started in 1610.

The 22-year average value of the modulation function (Φ_{CL80}) for the period 850–2000 AD is ~300 MeV, while the present-day average (1954–2006) is ~700 MeV. That is, the cosmogenic radionuclide data shows that the space era is not typical of the past 1,150 years. Figure 7.3-1 shows that the past 50 years and the Medieval Maximum define the upper limit of activity experienced during the past millennium, while the Grand Minima define the lower limit. In the next section, we

will see that this same degree of variability has existed throughout the past 10,000 years.

The almost complete absence of sunspots during the Maunder Minimum poses the important question – “do the solar 11 and 22-year sunspot cycles stop during a profound Grand Minima”? Does the solar dynamo stop? While the sunspot data cannot provide an answer, the cosmogenic radionuclide data with a temporal resolution of several years can. Figure 17.3.1-1 displays ^{10}Be data during the Spoerer Minimum (1420–1540 AD) and shows that the modulation function Φ_{CL80} varied by up to 200 MeV with an average period of ~ 11 years, and with indications of a 22-year periodicity as well. The ^{10}Be data from the Maunder Minimum and the ^{14}C data from both Grand Minima confirm this result (Stuiver and Braziunas 1993; Usoskin et al. 2001; Beer et al. 1998). Section 17.5 will show that the interplanetary magnetic field near Earth varied over the range 0.5–2.5 nT during the Spoerer Minimum, indicating that the solar magnetic fields continued to vary during this time. As discussed in Sect. 5.5, it is the strong magnetic fields in a sunspot that reduces the light output, thereby making it dark compared to the rest of the Sun. It seems plausible that while the cosmogenic radionuclide data show that the solar dynamo continued to operate during the Spoerer and Maunder Minima, the magnetic fields were not strong enough to produce sunspots that were visible by eye, or with the astronomical instruments of the late 1600s.

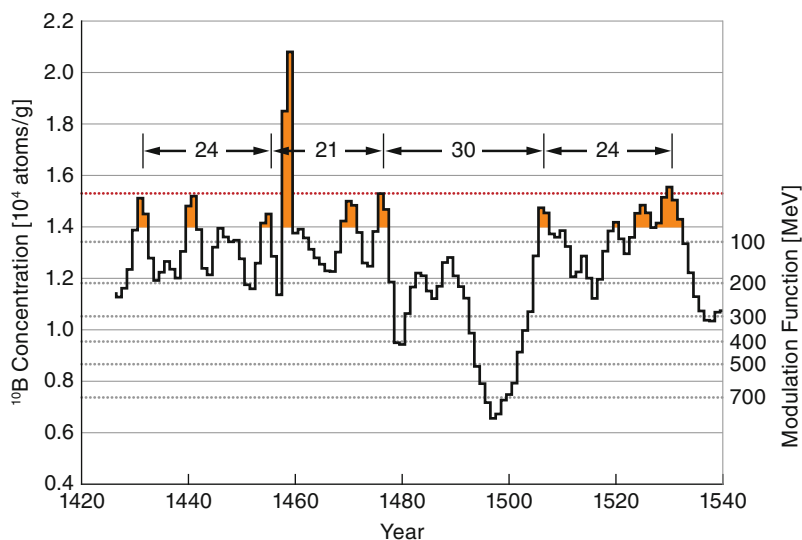


Fig. 17.3.1-1 The ^{10}Be data from Dye 3, Greenland, for the duration of the Spoerer Minimum (1420–1540 AD) (McCracken et al. 2004). The data points are shifted 1 year earlier to allow for the precipitation time of the ^{10}Be . A (1,4,6,4,1) moving weighted average (see Box 17.2.2.1) has been applied. The extreme points near 1460 AD are discussed in Sect. 18.3 as possibly due to a supernova or very large solar energetic particle event. The heavy red line is the estimated ^{10}Be concentration corresponding to no solar modulation. The modulation function, Φ_{CL80} , is given by the dotted lines and the scale on the right-hand side

17.3.2 *The Past 10,000 Years (the “Holocene”)*

The last period of major glaciation ended about 11,500 years ago. The concentration of ^{10}Be in ice halved abruptly at that time, because the annual precipitation of snow increased by a factor of about two (see Sect. 19.4). While corrections can be made for this to yield the flux of ^{10}Be into the polar archives, at the time of writing there has been little study of the temporal changes prior to the end of the ice age, other than for several particularly unusual periods (e.g. the “Laschamp event”—see Chap. 21). The ^{14}C record extends somewhat further into the past; however, we choose only to discuss the period for which both cosmogenic radionuclide data are available.

For completeness, we define some names. The glaciologists and geologists refer to the 11,500-year period after the end of glaciation as the “Holocene”. It is sometimes called the contemporary “interglacial”. This name reminds us that the Earth has been in an ice age for about 80% of the recent two million years, punctuated by interglacials roughly every 100,000 years. However, this does not mean that the climate was always cold during the ice ages, there being occasional short warm episodes, so-called Dansgaard–Oeschger events. The “Milankovitch cycles” responsible for the succession of glacial and interglacial epochs are due to long-term cyclic changes in the Earth’s orbit, and by precession and nutation of its rotation axis. The “little ice ages” such as occurred between 1400 and 1800 AD have a completely different origin, being primarily due to changes in the thermal output of the Sun (Sect. 17.5).

Figure 17.3.2-1 displays the production rates of ^{14}C and ^{10}Be since 9300 BP. It is clear that there are strong similarities between these completely independent records. Both show long-term changes that correlate with the estimates of the geomagnetic magnetic moment based on archaeometric studies (Sect. 21.2, also Fig. 5.8.1-2). Both show episodes of variations similar to those associated with the Spoerer, Maunder, and Dalton Minima discussed in the previous section (see also Fig. 7.3-1). The lowest panel presents the modulation function, Φ_{CL80} , derived from the data after removal of the geomagnetic effects, and after filtering to remove the effects of 11- and 22-year variations due to the Schwabe and Hale cycle solar variations.

There are several striking features of this figure, each emphasizing that solar activity has changed in a variable manner over this period. The Wolf and Spoerer Minima – the first two of the succession of four Grand Minima between 1300 and 1800 (seen in Fig 7.3-1) are clearly evident on the extreme right of Fig. 17.3.2-1 (recall that the cosmogenic radionuclide data are highest during a Grand (Solar) Minima). In the following, we refer to these four Grand Minima as the “Spoerer episode”. It is clear that such episodes were not the norm, e.g. the 1,300-year period 3000–4300 BP in which there were no Grand Minima greater than a Dalton Minimum. There was an episode of four Grand Minima in the vicinity of 5500 BP that was quite similar to the “Spoerer episode”; thus, the separations of the Grand Minima within the episodes were similar (~200 years), and the modulation function returned to similar low levels in each of the intensity peaks. There was a more restricted episode of two Grand Minima in the vicinity of 2500 BP; note that the duration of the earlier of these

Grand Minimum was similar to that of the Spoerer Minimum. There was also a prolonged episode of Grand Minima centred about 7300 BP.

The cosmogenic radionuclide data indicate that there were ~26 Grand Minima in the past 9,300 years. It is noticeable that the majority of them were of a similar duration to the Maunder Minimum (~70 years). Perhaps six were longer (~100 years) and similar to the Spoerer Minimum. In all, they show that the Sun was in the “Grand Minimum” condition for $\leq 15\%$ of the time. The evidence therefore suggests that there is a physical process in or associated with the Sun that results in these relatively short-lived episodes of low sunspot activity, and by inference, low solar magnetic field strengths, high intensities of galactic cosmic radiation, and low values of the modulation function ($\Phi_{\text{CL80}} \sim 100\text{--}200$ MeV). For the remaining 85% of the time, the cosmogenic radionuclide data indicate that the Sun was active, solar magnetic fields were relatively strong, the intensity of the galactic cosmic radiation was low, and the modulation function was high ($\Phi_{\text{CL80}} \sim 500$ MeV).

In 1982, Chuck Sonnet employed power spectrum analysis to demonstrate that there was a ~2,000-year periodicity in the rather limited ^{14}C data available at that time. He and Paul Damon called this the “*Hallstatt cycle*” and as outlined in Sect. 17.2.2, Sonnet’s result was later validated using ^{10}Be data, however, little was known about this periodicity, or whether it was of solar or terrestrial origin, until recently. Figure 17.3.2-1 removes much of the mystery. When the estimated modulation function, Φ_{CL80} , is filtered with a 1,000-year running average (red curve), groups of deep minima occur at 400 BP, 3100 BP, 5300 BP, and 7300 BP. Thus, these deep minima clusters occur with a periodicity of $(7300-400)/3 = 2,300$ years. Clearly (from Fig 17.3.2-1) the deep minima in the 1,000-year low-pass curve are due to the clusters of Grand Minima evident in the top two panels. That is, the ~2,300-year periodicity first detected by Sonnet was a consequence of a 2,300-year recurrence of clusters of Grand Minima. The 1,000-year low-pass curve also has small variations with a period of ~1,100 years; these are presumably the origin of the 970-year periodicity listed in Table 17.2.2-1. In summary, the Sun, and its magnetic dynamo, has exhibited periodic variations with a period of ~2,300 years throughout the Holocene in addition to the shorter period variations evident in the 400-year sunspot record.

Figure 17.3.2-1 illustrates the manner in which solar activity has varied between the episodes of Grand Minima. The 60-year average modulation function for the modern era (1940–2000 AD) is ~670 MeV, while the figure shows that it has been as high as 800 MeV in the past, indicating equivalent and higher solar activity on a number of occasions in the past. Note also the intervals such as that between 3000 and 4300 BP, where, as discussed previously, there were no Grand Minima. Figure 17.3.2-1 shows that Φ_{CL80} was consistently high throughout this interval, the 60-year average seldom decreasing below 400 MeV. High values of the modulation function (average = 550 MeV) for the 800-year period centred on 1900 BP indicate an active Sun similar to the modern day Sun. The absence of Grand Minima, and the consistently high values of Φ_{CL80} , indicate that the Sun was active during the intervals between the “Spoerer Episodes” throughout the Holocene.

Figure 17.3.2-1 suggests that the Hallstatt cycle may have been superimposed on a long-term increase in the modulation function between ~ 7000 and ~ 2000 BP. Comparison with other long cosmogenic records show similar, but not identical, increases over this period. As discussed in Chap. 13 there may be variable system (atmospheric) effects in the ^{14}C data, and ^{10}Be concentrations will be affected by long-term changes in snowfall due to climate change. It is possible that some or all of these very long-term changes are due to such atmospheric and climatic effects. A few new Holocene length ^{10}Be records are being measured at the time of writing, and together with those in Fig. 17.3.2-1, their use in principal components analysis should allow the existence, and properties of any long-term variations to be determined with greater accuracy.

Should they be verified, these long-term variations might be due to a number of different causes. Among them are (1) a long-term periodicity ($\sim 10,000$ year) in solar activity, (2) changes in the geomagnetic field at variance with the current archaeometric record (e.g. Fig. 5.8.1-2), and (3) long-term changes in the local interstellar spectrum of the galactic cosmic radiation (Sect. 5.7.2). All three would be of considerable intrinsic interest and the validation of the long-term changes and the identification of their cause are important tasks for the future.

17.3.3 *The Long Solar Minimum of 2007–2009*

As this book was nearing completion, the Sun entered a phase of lower activity. This was not unexpected, as we have seen there have been periods of reduced solar activity roughly every 100 years since the commencement of the sunspot record (Figs. 5.5-2 and 6.6-2). The previous section has shown that there have been 26 or more Grand Minima in the past 10,000 years. Table 17.2.2-1 has given the periodicities evident in the data. It was therefore only a matter of time before the period of high solar activity that commenced ~ 1945 came to an end. Using statistical methods, Abreu et al. (2008) predicted that it would occur within about a decade.

In late 2006, the sunspot number reached values similar to those attained during the previous four solar minima (~ 10). Solar activity continued to decline and by 2009 there were long periods without any sunspots. Figure 17.3.3-1 displays the concurrent changes in the cosmic ray intensities, the HMF, and the tilt angle of the heliospheric current sheet (Fig. 5.6-2). To emphasize the changes in behaviour, the figure also plots the same parameters for the previous $qA < 0$ solar minimum in 1986. All five panels in Fig. 17.3.3-1 demonstrate major changes in behaviour compared to the past 65 years of solar activity. Based on previous behaviour, sunspot numbers, HMF strength, and the heliospheric tilt angle would have been expected to start rising rapidly in 2007 while the cosmic ray intensities started to decrease. None of these changes happened until early 2010, by which time the sunspot cycle would have reached its maximum based on previous cycles. Solar activity has remained low in 2010 and it appears likely that the maximum sunspot

number for the 24th cycle will be very low compared to the cycles since 1946. Based on the worldwide network of neutron monitors, it is estimated that the modulation potential decreased to ~ 300 MeV in 2009. Using Fig. 6.6-1, the neutron monitor data predict that the sunspot minimum ^{10}Be concentration increased by 8% compared to the sunspot minima since 1954.

It is beyond the scope of this book to further discuss the details of this long sunspot minimum. It suffices to make several general comments about its significance in the study and use of the paleo-cosmic ray record.

- The data in Fig. 17.3.3-1, and other similar data, will provide the ability to improve our inter-calibration of the contemporary neutron monitor and satellite data with the cosmogenic radionuclide data in our archives.

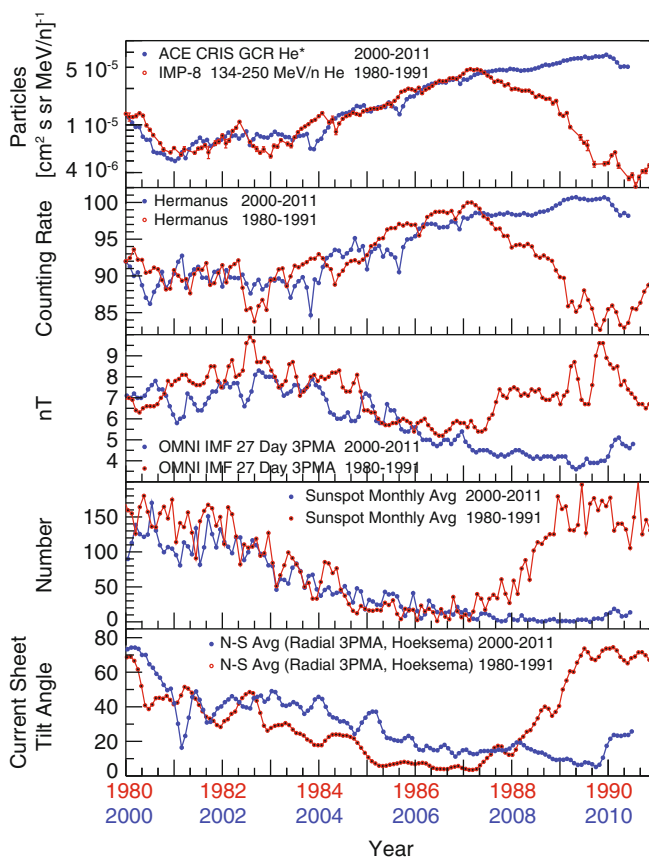


Fig. 17.3.3-1 Selected heliospheric properties during the long sunspot minimum 2007–2009, and comparison with the sunspot minimum 20 years previously (McDonald et al. 2010). *Panel 1:* Satellite observations of 134–250 MeV/nucleon He cosmic rays; *Panel 2:* The Hermanus neutron monitor; *Panel 3:* Satellite observations of the heliospheric magnetic field near Earth; *Panel 4:* Sunspot number; *Panel 5:* The tilt of the heliospheric current sheet

- Based on the cosmogenic and sunspot records, it appears possible that the next (25th) sunspot cycle may also be relatively inactive. As a consequence the HMF may be even lower still, and the cosmic ray intensity higher, further improving our ability to inter-calibrate the paleo-cosmic ray record to the present. Further, it would also improve our ability to use the paleo-cosmic ray data to estimate the HMF and TSI in the past (next two sections).

17.4 The Heliomagnetic Field Throughout the Past 10,000 Years

Spacecraft measurements of the HMF near the orbit of Earth commenced in 1964, and there is an essentially continuous record ever since (Fig. 17.4-1). It shows an 11-year variation of amplitude 4–5 nT, sitting on a baseline of ~5.2 nT. As discussed previously, the Sun has been very active throughout this whole period, and it is natural to speculate whether the baseline field strength was lower during the Maunder Minimum, or during the period of lower solar activity, circa 1900. If so, what would that imply about the magnetic fields of sunspots and of the Sun itself during those and earlier times? The cosmogenic radionuclide data provide the only way to investigate that prior to the commencement of routine sunspot observations in 1610.

As discussed in Sect. 5.5, sunspots represent large areas of very strong magnetic fields. The 11-year cycle and the long-term secular changes in the peak sunspot number (Fig. 5.5-2) all indicate that the solar magnetic fields change greatly with

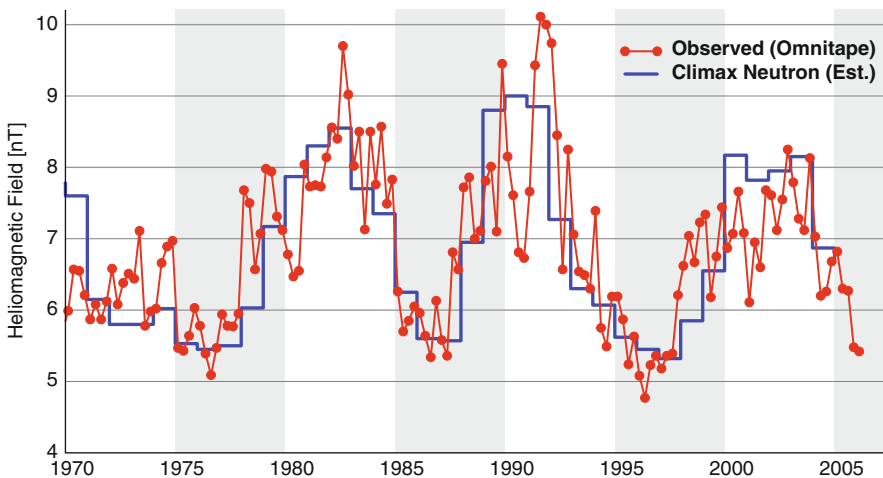


Fig. 17.4-1 Satellite measurements (*red*) of the strength of the heliospheric magnetic field (HMF) near Earth, and estimates (*blue*) obtained by using the cosmic ray propagation equation to invert the neutron monitor data from Climax, Colorado (McCracken 2007)

time. It is postulated that the sunspot and other solar magnetic fields are generated by an electromagnetic dynamo driven by the rotational motion of the Sun. Theoretical studies based on well-known magneto-hydrodynamic processes have been able to reproduce the 11-year cycle and the 22-year Hale cycle (e.g. Schrijver et al. 2002; Wang et al. 2002a). In summary, magnetic fields and magnetic processes are believed to play a key role in the short- and long-term changes in solar activity described in Sect. 5.5, and elsewhere in this book. However, much remains to be understood.

The interest in the solar magnetic fields is not purely academic. They play an important role in a number of phenomena that have serious implications for our technological society. Increased radiation damage to satellites, the disruption of radio services, and triggering geomagnetic storms that cause disruption of power distribution systems to name three which all have major economic implications. The solar magnetic fields also appear to contribute to the total solar irradiance (TSI), a key parameter in the study of climate change (Sect. 17.5).

For both practical and academic reasons then, there is a great interest in understanding the solar magnetic processes and forecasting them in advance. Using the Zeeman effect, occasional measurements of the sunspot and solar magnetic fields were commenced prior to 1900, however routine “maps” of the solar fields (helio-magnetographs) were not introduced until the 1960s. As outlined earlier, direct measurements of the interplanetary magnetic fields only started in 1964. Fifty years of data are totally inadequate to investigate processes that may have time scales >10,000 years. The sunspot data starting in 1610 are an improvement, but very limited. They give no indication of the polarity of the magnetic cycles in the past, and whether the phase of the Hale cycle changes going through a Grand Minimum, say. Nor does it tell us if the solar field disappeared completely during the most profound Grand Minima such as the Spörer Minimum.

Modern theories and observations indicate that the strength of the HMF is strongly influenced by the strength of the solar fields (Solanki et al. 2002; Wang et al. 2002b). As discussed in Sect. 5.7, the modulation of the cosmic radiation is a strong function of the strength of the HMF. Using the cosmic ray transport equation to “work backwards”, a method was developed in 2004 to compute the time-dependent strength of the HMF that yields the observed variations in the cosmogenic ^{10}Be data (Caballero-Lopez et al. 2004). The field strength also largely determines the values of the diffusion coefficient in the definition of the modulation function (Eq. 5.7.3-2), and this led to alternative methods to invert the cosmogenic radionuclide data (McCracken 2007; Steinhilber et al. 2010). Estimates based on this inversion methodology are displayed in Fig. 17.4-1 (using modern neutron monitor data) and in Figs. 17.4-3 and 17.4-4 (using cosmogenic ^{10}Be).

That is, the cosmic ray intensity at Earth can be regarded as the output of a form of interplanetary magnetometer. The cosmogenic radionuclide data are therefore a permanent record of the output of that interplanetary magnetometer. At the present time, they are the only known way to estimate the properties of the solar and HMF prior to 1610.

Before using the cosmogenic radionuclide data as a “cosmic magnetometer” it is clearly important to check them against other techniques that predict or estimate the HMF between 1610 and 1965. Two others techniques exist:

1. Forward models, that use the known magnetic properties of sunspots, magneto-hydrodynamic theory, and the historical sunspot record, to compute an estimate of the HMF near Earth since 1610. The output of one of these models (Solanki et al. 2002) is shown in Fig. 17.4-2; it is consistent with other forward models. These models all involve ad hoc factors, which remain to be established by independent means. They all estimate that the 11-year average HMF had a strength of $\sim 1\text{--}2$ nT at the end of the Maunder Minimum, varying subsequently in a manner that reflects the Gleissberg cycle of solar activity.
2. Two methods based upon the manner in which the magnetic pressure in the solar wind introduces recognizable effects into the geomagnetic record that extends back to 1872. These two methods infer divergent estimates for the change in the HMF since the Gleissberg Minimum of 1889–1902 as is evident in Fig. 17.4-2. While there are differences between the several results, they all agree that (1) the HMF has exhibited a long-term increase since 1700, and since 1900, and (2) there were large decreases during the Maunder and Dalton Minima. There is reasonable agreement between two of the estimates and the results based on the

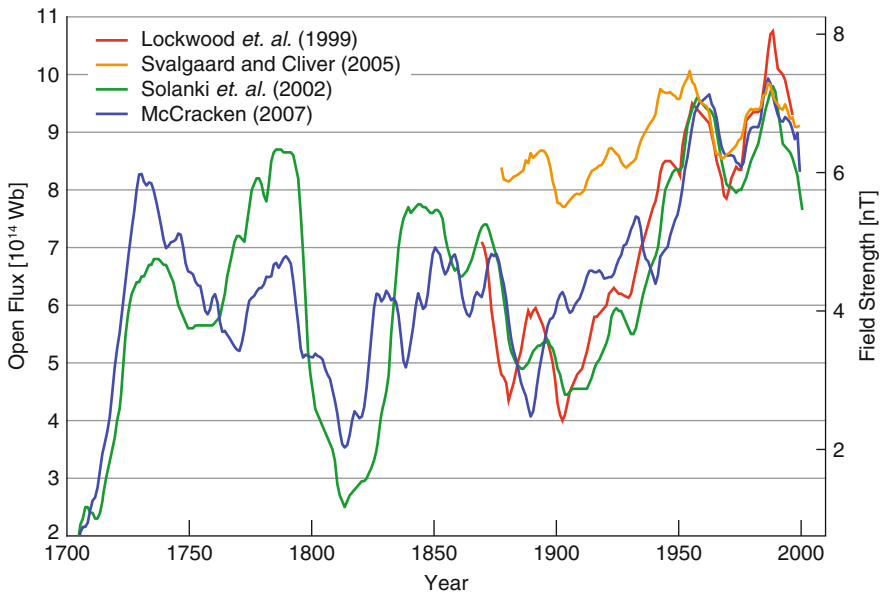


Fig. 17.4-2 Comparison of estimates of the interplanetary magnetic field: two estimates based on the measured properties of the geomagnetic field (Lockwood et al. 1999; Svalgaard and Cliver 2005), a forward model based on sunspot number (Solanki et al. 2002), and an estimate obtained from the cosmogenic ^{10}Be as outlined in the text (McCracken 2007)

cosmogenic radionuclide data, providing confidence to use them to investigate the properties of the solar magnetic fields in the past.

Based on the annual data record from Dye 3 in Greenland, Fig. 17.4-3 displays the annual estimates of the HMF near Earth since 1428. This shows that the amplitude of the 11-year cycle in the HMF remained in the range 2.0–3.0 nT throughout the whole period. It further shows that the HMF field at sunspot minimum was <1.0 nT during the Spoerer Minimum, rising steadily since the Maunder Minimum to ~5.2 nT during the sunspot minima since 1954. Figure 17.4-1 shows that there was an 11-year variation in the HMF of amplitude ~4.0 nT during the period of high solar activity since 1965; it is remarkable that the amplitude was comparable (~2.0 nT) during the most profound portion of the Spoerer Minimum, 1420–1485. As discussed in Sect. 17.3.1, this provides strong evidence that the 11-year solar cycle, and the solar dynamo continues throughout a deep Grand Minimum.

The question remains – “does the phase of the Hale magnetic cycle change during a Grand Minimum”? The cosmogenic radionuclide data provide the potential to answer this question in the future. Thus, Sect. 5.7.5 has discussed the fact that the polarity of the HMF determines the shape of the 11-year cycle in the cosmic ray

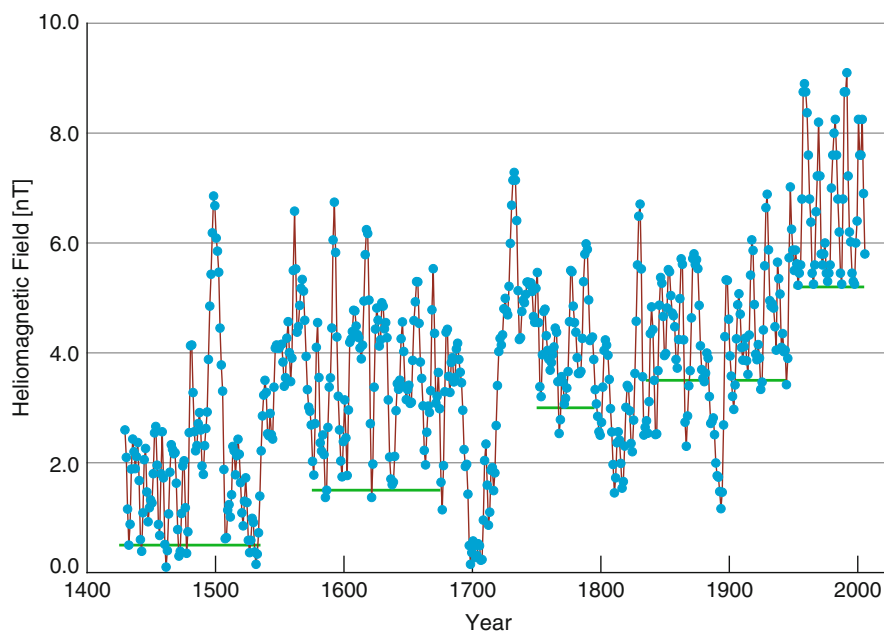


Fig. 17.4-3 The estimated strength of the HMF since the fifteenth century, obtained by inversion of the cosmogenic ^{10}Be data from Dye 3, Greenland. These are annual data, smoothed with a (1,4,6,4,1) binomial filter. The *green lines* indicate baseline values that appear to have changed between the several Grand Minima. The high frequency oscillations are due to the 11-year cycles of solar activity

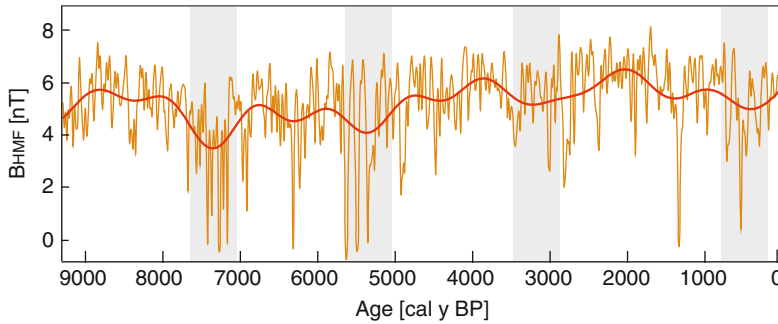


Fig. 17.4-4 The estimated strength of the HMF at Earth for the past 9,300 years. The data are averaged with 40-year and 1,000-year low pass filters (Steinhilber et al. 2010)

intensity, resulting in a succession of broad and sharp maxima of the cosmic ray intensity. Thus for $qA > 0$, the peak is broad, for $qA < 0$ it is sharp. This effect is large enough to be seen in the cosmogenic radionuclide data, once there are several independent annual records that will allow the statistical noise to be reduced using principal component analysis or some similar mathematical procedure.

Using ^{10}Be data from the GRIP ice core from Greenland, Fig. 17.4-4 presents estimates of the strength of the HMF for the past 9,300 years. The period since the Spoerer Minimum (Fig. 17.4-3) is at the extreme right. From Fig. 17.4-4, it is clear that the 40-year average HMF has varied over the range 1–8.5 nT throughout the Holocene. The red curve presents the HMF after being averaged in a 1,000-year low-pass filter, and shows that there were Hallstatt (~2,300 years) and Eddy (~970 years) cycles in the HMF of amplitude ~2 nT.

Finally, we sound a word of caution as befits a scientific discipline that it is still evolving as more data are obtained. While there is good general agreement between cosmogenic ^{10}Be from different parts of the world (e.g. Fig. 7.3-2), there are occasional relatively large differences in the long-term trends (Webber and Higbie 2010), and also over shorter (~50 years) periods (McCracken et al. 2004). These differences are probably due to localized climate effects. In the future, the use of principal components analysis to combine several independent ^{10}Be records should reduce the consequent residual errors in the estimates of the HMF to a considerable degree.

17.5 Solar Irradiance and Terrestrial Climate

The effects of climate change upon many facets of the environment, and on human life, have become questions of great importance over the past 20 years. The role of the anthropogenic gasses (CO_2 , CH_4 , etc.) has been documented in detail (IPCC 2007), and considerable attention given to the role of a variable output from the Sun as well.

Accurate satellite measurements of the TSI commenced about 1978 as shown in Fig. 17.5-1, and it is clear that it has varied by about 1.4 W m^{-2} in phase with the

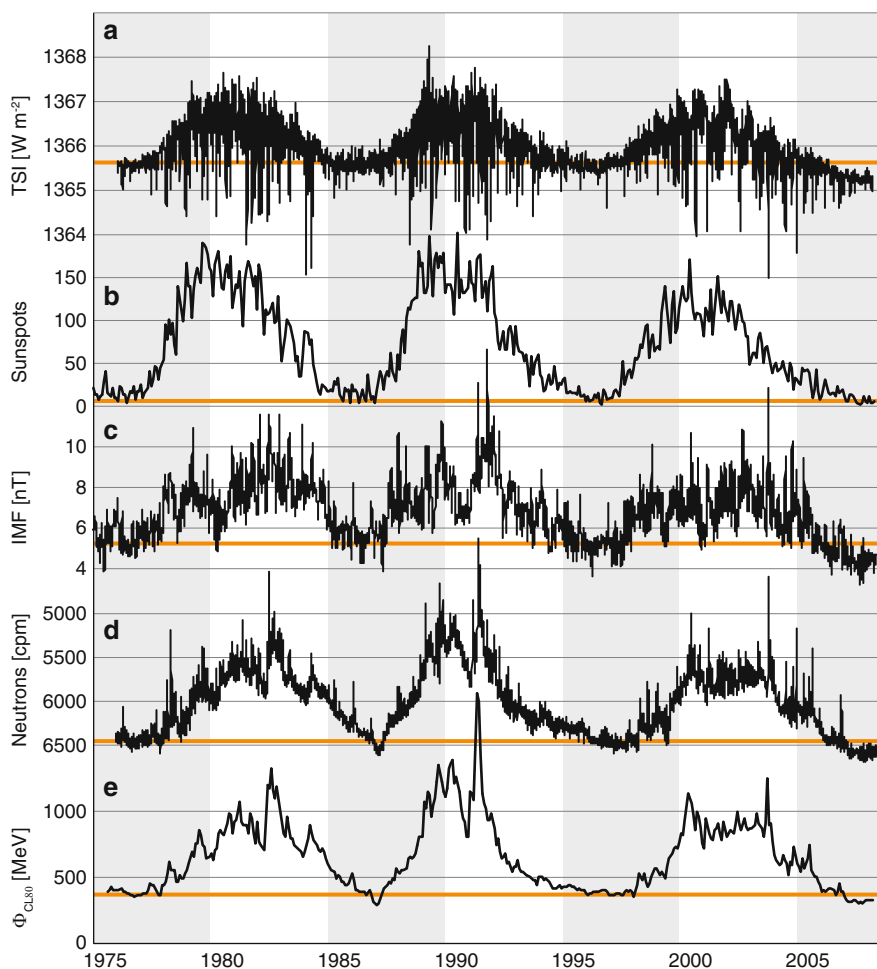


Fig. 17.5-1 The relationships between the total solar irradiance (TSI) and other solar and solar system properties. *Panel A:* the TSI; *Panel B:* the sunspot number; *Panel C:* the field strength of the interplanetary magnetic field (IMF); *Panel D:* Oulu neutron monitor counting rate (*inverted*); and *Panel E:* the modulation function, Φ_{CL80} . Note the extended sunspot minimum at the end of solar cycle 23 in 2008, as discussed in Sect. 17.3.3

solar cycle. While the sunspots are less bright than the ambient photosphere, the integrated effects of the “faculae” and “the magnetic network” over-correct for this, leading to a higher TSI when sunspot numbers are high. It should be noted that all three of these factors – the sunspots, the faculae, and the magnetic network – are direct consequences of the magnetic fields of the Sun. The $\sim 0.1\%$ variation in the TSI throughout the 11-year cycle has a small but discernible effect upon the Earth’s climate which is difficult to quantify (Beer et al. 2000; IPCC 2007). Thus, in

addition to the direct heating, there are positive feedback effects that may amplify the solar effects by a factor of two or more (Meehl et al. 2009).

It has long been recognized that the climate of some parts of the Earth suffered low temperatures – called “the little ice ages” – at the time of the Spoerer, Maunder, and Dalton Grand Minima in sunspot numbers. Clearly, we would like to know the value of the TSI at those times, to determine whether the little ice ages were causally connected to the reduced solar activity. Unfortunately, the Sun has remained in a strongly active state for the whole of the space age, and until recently the satellite data have not been able to provide any insight into whether there is a long-term secular trend in the TSI as a function of the strength of the solar cycle.

The cosmogenic radionuclide data provide one of the few sources of information that allow us to investigate the solar-climate connection in the past, both in a qualitative and a quantitative manner. They also provide the means to estimate the TSI in the past (Steinhilber et al. 2009). Two examples of the correlation between solar activity, and climate, are shown in Figs. 17.5-2 and 17.5-3. The first displays the retreating and advancing phases of the Aletsch Glacier in Switzerland, as determined from carbon dating of the trees killed by the advancing ice (Holzhauser et al. 2005). The modulation function (Φ_{CL80}) derived from the

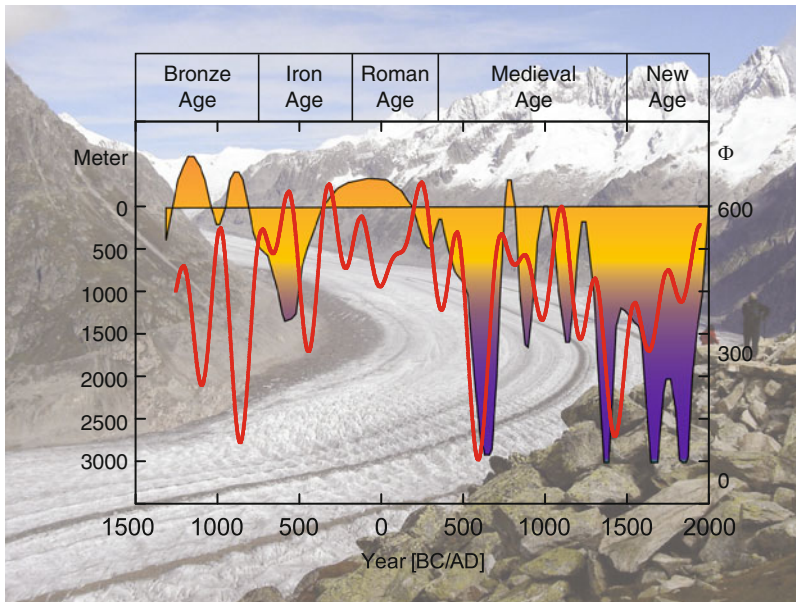


Fig. 17.5-2 Comparison of the extension of the “great Aletsch glacier” in the Swiss Alps (*black*) and the modulation function (*red*). Photographic records show that the glacier has retreated by more than 3 km since the nineteenth century. However, the present retreat distance is not unique; similar retreats occurred in Medieval and Roman times (Holzhauser et al. 2005). The changes in the modulation function show that large extensions of the glacier in the past occurred at times of low solar activity. The data on glacier extent become increasingly uncertain before ~100 AD. The glacier is seen in the background image

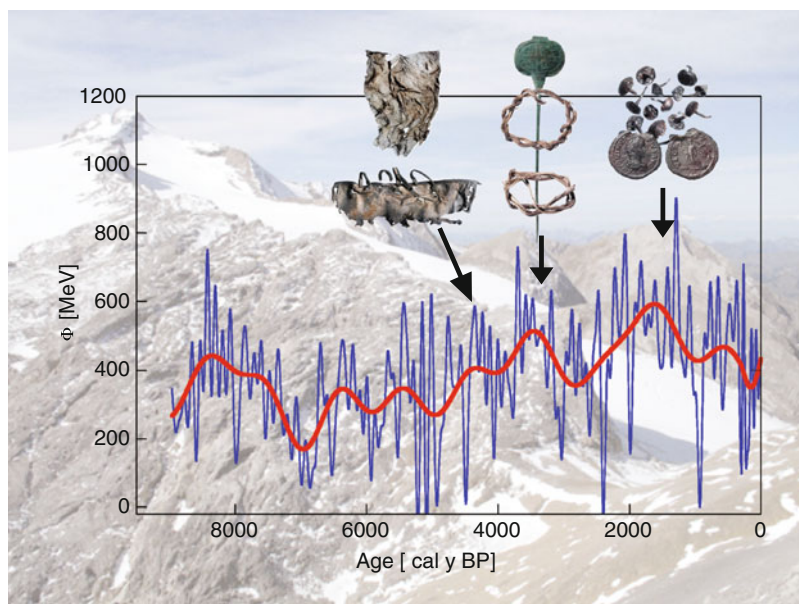


Fig. 17.5-3 Archaeological finds indicating warm periods in the European Alps, compared to the modulation function (Φ_{CL80}) derived from the cosmogenic record. The artefacts were ~4500 BP – pieces of trousers and shoes made of goat leather; ~3500 BP – bronze needle and wooden jewellery; ~1900 BP – Roman coins, shoe nails, etc. (Grosjean et al. 2007). The mountain pass where these artifacts were found is seen in the background image

cosmogenic radionuclide data is superimposed. It is clear that the glacier advanced when Φ_{CL80} (and solar activity) was low during the Dalton, Maunder, and Spörer Minimum. Note, in addition, the good match between the Grand Minimum of 600 AD, and the advance of the glacier, long before any changes in the concentration of anthropogenic gases in the atmosphere.

Over the past several warm decades, the retreating ice in the high European mountains has revealed evidence that there have been similar warm periods in the past. Some of that evidence is in the form of artefacts left behind by travellers passing along the trade routes over the high mountain pass “Schnidejoch” (Fig. 17.5-3). The artefacts were found in places that have been ice covered throughout the twentieth century. Using carbon dating, the ages of the artefacts were found to be in small clusters as shown in the figure. This indicates that the trade routes were only open for limited periods. The figure shows that the clusters coincide with high values of the modulation function (Φ_{CL80}) similar to those of the present epoch. That is, the mountain passes that are open now (active Sun) were also open in Roman and Neolithic times when the totally independent cosmogenic radionuclide data shows that the Sun approximated the present day level of activity.

The climate system is very complex, with various time lags, time constants, various feedback mechanisms, etc. Their intertwined effects can only be understood and quantified for the future using complex mathematical models. Those models use “forcing functions” to quantify the various physical factors that may influence the climate, and the TSI is the one used to input the “solar forcing” effects. As noted earlier, while we have instrumental TSI data since 1978, the satellite data have provided no insight into long-term trends and, in particular, to the value of TSI during the Maunder and other Grand Minima. In some studies, the cosmogenic ^{10}Be or ^{14}C data were used as “proxies” for the solar input (e.g. IPCC 2007). Since the physics behind changes in total and spectral solar irradiance is not yet understood it is difficult to derive quantitative records of solar forcing. This leads to rather large discrepancies between different reconstructions. More details on the solar influence on climate change can be found in two review articles (Gray et al. 2010; Wanner et al. 2008).

It was noted earlier that the satellite measurements show that all of the short-term (11-year cycle) changes in the TSI are associated with variations in the magnetic properties of the Sun. Examination of Fig. 17.5-1 provides a tantalizing suggestion that during the period 2007–2010 there may have been long-term reduction in TSI due to the long-term reduction in the solar magnetic fields. Thus, note that both the HMF and TSI exhibited 11-year fluctuations above the baseline values of 5.2 nT and $1,365.6 \text{ Wm}^{-2}$, respectively, for the three sunspot cycles between 1975 and 2006. They suggest a TSI sensitivity of $0.4 \text{ Wm}^{-2} \text{ nT}^{-1}$. Having decreased to the baseline values by early 2006, both the HMF and TSI continued to decrease until by mid-2008 they were $\sim 1 \text{ nT}$ and $\sim 0.4 \text{ Wm}^{-2}$ below the previous baseline value. Other evidence suggests that the Sun may be entering a period of reduced activity (Sect. 17.3.3). If verified, this suggests that there will be a long-term change in the TSI that is closely related to the strength of the HMF (Abreu et al. 2008; Steinhilber et al. 2009).

If this relationship is verified, the cosmogenic radionuclide data will provide a source of quantitative information regarding the values of TSI in the past. As summarized in earlier sections of this book, we now have (a) inter-calibrated cosmic ray (Sect. 6.6) and modulation function records for the whole Holocene, up to the present day; and (b) methodologies that have used those inter-calibrated records to estimate the strength of the HMF for that whole period (Sect. 17.4). If the relationship between TSI and HMF is validated, this will allow the TSI to be estimated for any time in the past 10,000 years. Furthermore, the modulation function is now known in real time based on the worldwide network of neutron monitors, allowing instant comparison with anytime in the past, if desired. Further still, several new ^{10}Be data sets are being analyzed at present, promising an improved signal-to-noise ratio for the solar signal in the near future. In sum, it seems likely that over the next 5 years the paleo-cosmic ray data will provide greatly improved estimates of the long-term secular changes in TSI for input into climate change models.

17.6 Radiation Doses on Earth and in Space in the Future

Section 17.3 has described how the galactic cosmic radiation has varied over the past 10,000 years. Chapter 8 has discussed the information gained from the nitrate data regarding the generation of cosmic rays by the Sun in the past. Contrary to intuition, both sets of experimental evidence indicate that lower solar activity results in higher radiation dose rates in the atmosphere and in space. As discussed in Sect. 17.3, the paleo-cosmic ray data show that the “space age” represents a period of relatively high solar activity and low cosmic ray intensity. Thus only 100 years ago, during several decades of low solar activity, the intensity of the GCR was substantially higher (Fig. 6.6-2), and the frequency of occurrence of large solar energetic particle (SEP) events was up to a factor of six greater than during the past 50 years (Fig. 8.2.3-2).

Based on the sunspot and cosmogenic records, it appears possible that the period of high solar activity which began in the middle of the twentieth century is coming to an end (Abreu et al. 2008). As outlined in the preceding paragraph this would lead to higher radiation dose rates, and to more rapid degradation of communication and other satellite systems, and higher radiation doses to airline crew and passengers. Our ability to estimate the severity of these threats is almost totally dependent on the cosmogenic and nitrate data discussed previously. The probability of high SEP fluencies is shown in Fig. 8.3.3-2. The ^{10}Be and ^{14}C data provide the only means to estimate the variability of the galactic cosmic ray intensities in the past and the attendant changes to the spectrum. Thus, the spectrum and intensities at any time in the past 10,000 years can be estimated using equation 5.7.3-1 and the value of the modulation function for the epoch of interest (read from Figs. 7.3-1, 17.2.1-2 and 17.3.2-1). It is beyond the scope of this book to discuss these matters in any detail, other than to point out that, for the first time, the existing cosmogenic radionuclide data provides the means to determine the extent to which the future may differ from the past 50 years.

17.7 Quantitative Measures of Solar Activity for the Past

Apart from its intrinsic interest, solar activity has substantial impacts on a large number of our modern activities. For example, as outlined in Sect. 17.4, solar activity disrupts radio services, triggers geomagnetic storms, and interferes with mineral exploration on a day-to-day basis. For this reason, the sunspot number is an important predictive parameter in modern society. Estimates of what it was in the past are useful in planning for the future.

As we saw in Figs. 17.5-2 and 17.5-3, changes in climate in the past appear to have been correlated with solar activity derived from the cosmogenic record. In Sects. 17.5 and 17.6, we have seen that the TSI and radiation dose in the past can also be inferred from the cosmogenic record. It is reasonable to speculate that many climate-related sciences – biology, palaeontology, archaeology, etc. – will

find the cosmogenic record useful in interpreting their observations of the recent past.

It is therefore necessary that we have a parameter that quantifies solar activity that extends throughout the Holocene up to the present day. One such could be the estimated neutron monitor counting rate derived from the cosmogenic radionuclide data, as shown in Fig. 6.6-2 for the interval 1428–2005. Some studies have used the ^{10}Be and ^{14}C data themselves for this purpose, however such records are hard to compare between themselves and with the present.

At the present there are two parameters that appear the most suitable for future use. They are (1) reconstructed sunspot numbers, and (2) the modulation function. We discuss both briefly in the following.

17.7.1 Reconstructed Sunspot Numbers

Figure 17.7.1-1 presents a reconstruction based on the ^{14}C record (Solanki et al. 2004). Conceptually, this reconstruction is based on the linking together of three separate mathematical models: (a) the dependence of the solar magnetic flux on sunspot number, (b) the modulation of the galactic cosmic radiation by the HMF, and (c) the production of the cosmogenic radionuclides in the Earth's atmosphere and their sequestration in polar ice or tree rings. Inverting the process, the sunspot number is estimated corresponding to the observed cosmogenic radionuclide data. Allowance has been made for the effects of the changing geomagnetic field. Note in particular that

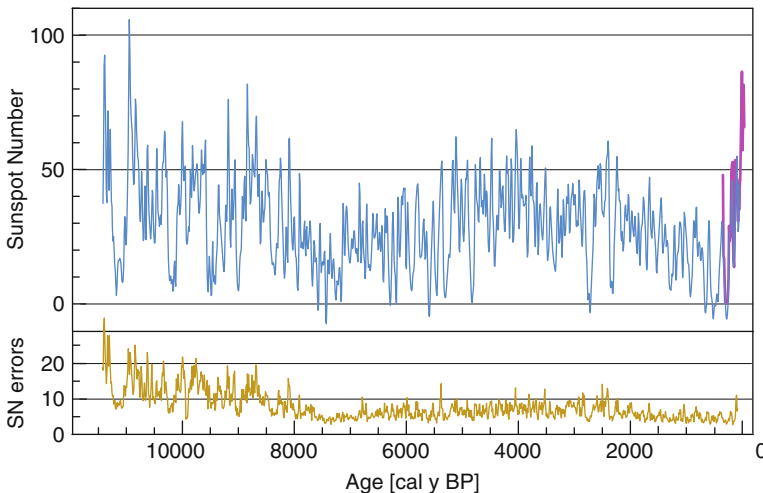


Fig. 17.7.1-1 Reconstructed sunspot number from 11,500 BP to the present after (Solanki et al. 2004). *Top panel:* 10-year average sunspot number reconstructed from the $\Delta^{14}\text{C}$ record (blue line) and 10-year averaged group sunspot numbers obtained from the telescopic record since 1610 (red line). *Bottom panel:* Standard deviations of the uncertainties in the reconstructed sunspot numbers

these are 10 year average sunspot numbers – thus the 19th sunspot cycle 1954–1965 which had a peak monthly sunspot number of ~ 180 has a peak 10-year average of ~ 87 in the figure. The bottom panel gives the uncertainties in the reconstruction.

Based on their reconstruction, Solanki et al. (2004) conclude that solar activity has been very high for the past 70 years, the previous period of equally high solar activity having occurred more than 8,000 years ago.

17.7.2 Modulation Function

The modulation function, Φ_{CL80} , was discussed in detail in Sects. 5.7.3 and 5.7.4 and has appeared in many figures throughout this book. In summary, it is a parameter that quantifies the effects of solar activity upon the cosmic ray spectrum in the vicinity of Earth. As such it is a convenient parametric description of solar activity. In concept, its derivation is similar to the reconstruction process used in Sect. 17.7.1, except that it does not include the mathematical model that links the heliomagnetic flux to sunspot number.

The modulation function has the great advantage that it allows estimates to be made of the cosmic ray spectrum at any time in the past. Given a tabulation of the modulation function Eq. 5.7.3-1 can be used to compute the intensity of any desired cosmic ray energy, or the spectrum, at any time in the past.

In Sect. 5.7.3, we have discussed the uncertainties that exist in the definition and computation of the modulation function (similar uncertainties exist in the case of the reconstructed sunspot number). The reader is directed to Sect. 5.7.4 where we have stated the convention used throughout this book, and in particular, in all the figures. Briefly, the cosmogenic and other data have been used to estimate the modulation function Φ_{CL80} for protons using (a) the Castagnoli and Lal (1980) local interstellar spectrum (LIS) and (b) the Gleeson and Axford (1968) formulation of the modulation function, and taking into account the proton, Helium, and heavy components of the cosmic radiation. As discussed in Sect. 5.7.3, the use of different formulations of the LIS will result in different values of Φ . Figure 5.7.3-3 can be used to convert between values of Φ obtained with other LIS.

References

- Abreu JA, Beer J, Steinhilber F, Tobias SM, Weiss NO (2008) For how long will the current grand maximum of solar activity persist? *Geophys Res Lett* 35:L20189
- Bard EG, Raisbeck M, Yiou F, Jouzel J (1997) Solar modulation of cosmogenic nuclide production over the last millennium: comparison between C-14 and Be-10 records. *Earth and Planetary Science Letters* 150(3-4):453-462
- Beer J, McCracken KG (2007) Evidence for solar forcing: some selected aspects. Paper presented at Climate and Weather of the Sun-Earth System (CAWSES), TERRAPUB, Tokyo, 2009, Kyoto

- Beer J, Tobias SM, Weiss NO (1998) An active Sun throughout the Maunder minimum. *Solar Phys* 181:237–249
- Beer J, Mende W, Stellmacher R (2000) The role of the Sun in climate forcing. *Quat Sci Rev* 19(1–5):403–415
- Beer J, McCracken KG, Abreu JA, Heikkilä U, Steinhilber F (2007) Long-term changes in cosmic rays derived from cosmogenic radionuclides. Paper presented at 30th International Cosmic Ray Conference ICRC07, Merida
- Caballero-Lopez RA, Moraal H, McCracken KG, McDonald FB (2004) The heliospheric magnetic field from 850 to 2000 AD inferred from/sup 10/Be records. *J Geophys Res* 109:A12102
- Castagnoli G, Lal D (1980) Solar modulation effects in terrestrial production of C-14. *Radiocarbon* 22(2):133–158
- Damon PE, Jirikowic JL (1992) The Sun as a low-frequency harmonic oscillator. *Radiocarbon* 34:199–205
- Damon PE, Sonett CP (1991) Solar and terrestrial components of the atmospheric ¹⁴C variation spectrum. In: Sonett CP, Giampapa MS, Matthews MS (eds) *The Sun in Time*. University of Arizona press, Tucson, pp 360–388
- Eddy JA (1976) The Maunder minimum. *Science* 192(4245):1189–1201
- Gleeson LJ, Axford WI (1968) Solar modulation of galactic cosmic rays. *Astrophys J* 154:1011–1026
- Gray LJ et al (2010) Solar influences on climate. *Rev Geophys* 48(4):RG4001
- Grosjean M, Suter PJ, Trachsel M, Wanner H (2007) Ice-borne prehistoric finds in the Swiss Alps reflect Holocene glacier fluctuations. *J Quat Sci* 22(3):203–207
- Holzhauser H, Magny M, Zumbuhl HJ (2005) Glacier and lake-level variations in west-central Europe over the last 3500 years. *Holocene* 15(6):789–801
- IPCC (2007) Fourth Assessment on Climate Change – The Physical Science Basis Rep
- Lal D, Peters B (1962) Cosmic ray produced isotopes and their application to problems in geophysics. In: Wilson JG, Wouthuysen SA (eds) *Progress in elementary particle and cosmic ray physics*. North-Holland publishing company, Amsterdam
- Lockwood M, Stamper R, Wild MN (1999) A doubling of the Sun's coronal magnetic field during the past 100 years. *Nature* 399(6735):437–439
- McCracken KG (2007) Heliomagnetic field near Earth, 1428–2005. *J Geophys Res-Part A-Space Phys* 112:A09106, 09101–09109
- McCracken KG, McDonald FB, Beer J, Raisbeck G, Yiou F (2004) A phenomenological study of the long-term cosmic ray modulation, 850–1958 AD. *J Geophys Res-Space Phys* 109:A12103
- McCracken KG, Beer J, McDonald FB (2005) The long-term variability of the cosmic radiation intensity at Earth as recorded by the cosmogenic nuclides. In: Geiss J, Hultqvist B (eds) *The solar system and beyond, ten years of ISSI*. ESA Publication, The Netherlands, pp 83–98
- McDonald FB, Webber WR, Reames DV (2010) Unusual time histories of galactic and anomalous cosmic rays at 1 AU over the deep solar minimum of cycle 23/24. *Geophys Res Lett* 37:L18101
- Meehl GA, Arblaster JM, Matthes K, Sassi F, van Loon H (2009) Amplifying the Pacific climate system response to a small 11-year solar cycle forcing. *Science* 325(5944):1114–1118
- Muscheler R, Beer J, Wagner G, Laj C, Kissel C, Raisbeck GM, Yiou F, Kubik PW (2004) Changes in the carbon cycle during the last deglaciation as indicated by the comparison of Be-10 and C-14 records. *Earth Planet Sci Lett* 219(3–4):325–340
- Peristykh AN, Damon PE (2003) Persistence of the Gleissberg 88-year solar cycle over the last 12,000 years: evidence from cosmogenic isotopes. *J Geophys Res-Space Phys* 108(A1), art. no.-1003
- Reimer PJ et al (2004) IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–1058
- Schrijver CJ, DeRosa ML, Title AM (2002) What is missing from our understanding of long-term solar and heliospheric activity? *Astrophys J* 577(2):1006–1012
- Smart DF, Shea MA, McCracken KG (2006) The Carrington event: possible solar proton intensity-time profile. *Adv Space Res* 38(2):215–225

- Solanki SK, Schussler M, Fligge M (2002) Secular variation of the Sun's magnetic flux. *Astron Astrophys* 383(2):706–712
- Solanki SK, Usoskin IG, Kromer B, Schussler M, Beer J (2004) Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature* 431(7012):1084–1087
- Steinhilber F, Abreu JA, Beer J (2008) Solar modulation during the Holocene. *Astrophys Space Sci Trans* 4:1–6
- Steinhilber F, Beer J, Frohlich C (2009) Total solar irradiance during the Holocene. *Geophys Res Lett* 36:L19704
- Steinhilber F, Abreu JA, Beer J, McCracken KG (2010) Interplanetary magnetic field during the past 9300 years inferred from cosmogenic radionuclides. *J Geophys Res-Space Phys* 115:A01104
- Stuiver M, Braziunas TF (1993) Sun, ocean, climate and atmospheric $^{14}\text{CO}_2$, an evaluation of causal and spectral relationships. *Holocene* 3:289–305
- Stuiver M, Braziunas TF, Becker B, Kromer B (1991) Climatic, solar, oceanic, and geomagnetic influences on late-glacial and Holocene atmospheric C-14/C-12 change. *Quat Res* 35(1):1–24
- Suess HE (1970) The three causes of the secular C14 fluctuations, their amplitudes and time constants. In: Olsson IU (ed) *Radiocarbon variations and absolute chronology*. Almquist and Willsell, Stockholm, pp 595–605
- Svalgaard L, Cliver EW (2005) The IDV index: its derivation and use in inferring long-term variations of the interplanetary magnetic field strength. *J Geophys Res-Space Phys* 110: A12103
- Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79(1):61–78
- Usoskin I, Mursula K, Kovaltsov G (2001) Heliospheric modulation of cosmic rays and solar activity during the Maunder minimum. *J Geophys Res Space Phys* 106(A8):16039–16046
- Vonmoos M, Beer J, Muscheler R (2006) Large variations in Holocene solar activity: constraints from Be-10 in the Greenland Ice Core Project ice core. *J Geophys Res-Space Phys* 111: A10105
- Wang YM, Sheeley NR, Lean J (2002a) Meridional flow and the solar cycle variation of the Sun's open magnetic flux. *Astrophys J* 580(2):1188–1196
- Wang YM, Lean J, Sheeley NR (2002b) Role of a variable meridional flow in the secular evolution of the Sun's polar fields and open flux. *Astrophys J* 577(1):L53–L57
- Wanner H et al (2008) Mid- to late Holocene climate change: an overview. *Quat Sci Rev* 27(19–20):1791–1828
- Webber WR, Higbie PR (2010) What Voyager cosmic ray data in the outer heliosphere tells us about Be-10 production in the Earth's polar atmosphere in the recent past. *J Geophys Res-Space Phys* 115:05102

Chapter 18

Galactic Astronomy

18.1 Introduction

As discussed in Sect. 5.4, and in Chap. 17, the cosmogenic data provide us with information about the past. In particular, we have studied the variability of the Sun, and in Chap. 21 will study the secular changes in the geomagnetic field.

In the same manner, we can envisage ways in which the cosmogenic data will provide information about the galaxy in which we live. Note the word “envisage”. This chapter is the only one in this book that does not deal with observations and interpretations that are well tested and widely accepted by the scientific community. There is agreement regarding what might be seen in the cosmogenic and similar data, which is summarized in the following. There has been attribution of some observations with an astronomical origin, and these are outlined. In general, however, this chapter is about what we may see in the cosmogenic and other data in ice and other archives as they become more accurate, of longer time span, or simply looked at more closely in the future.

We outline the use of cosmogenic and other methods for three types of astronomical investigations. The first is where the concentrations of a stable and a radioactive cosmogenic nuclide are used to obtain the exposure age of meteorites and moon rock. This technique permits study of the history of the solar system over time scales of billions of years. In the second, we discuss how variations in the properties of interstellar space immediately outside the heliosphere will result in changes in the production of cosmogenic radionuclides over periods of the order of 30,000 years. In the third method, we seek to identify discrete events such as supernovae that resulted in an enhanced production of a radioactive nuclide with a long half-life, such as ^{36}Cl (0.301 Myr), ^{26}Al (0.73 Myr), and ^{10}Be (1.39 Myr). Allowing that they may provide useful information for up to five half-lives in the past limits our investigations of the past to ~ 5 Myr (for example, from meteoritic material or deep-sea sediments).

18.2 Galactic Structure

Astronomical observations (optical and radio) have suggested that our galaxy has four spiral arms as depicted in Fig. 18.2-1, where the dots represent an assumed distributions of the supernovae that are the sources of the galactic cosmic rays (Sect. 5.4). The interstellar magnetic fields severely inhibit the motion of the cosmic rays and initially restrict them to their spiral arm of birth. They slowly diffuse out into the space between the spiral arms.

Figure 18.2-2 illustrates a model based on this concept and presents data from several sources that have been discussed in association with this model. Panel (a) postulates the times when the solar system passed through the spiral arms of the galaxy.

Panel (b) is the result of a diffusion calculation that yields the galactic cosmic ray intensity near 1 GeV (i.e. near the peak in ^{10}Be response curve – Fig. 6.5-2) relative to the present day value (Busching and Potgieter 2008). The model shows that the intensity outside the spiral arms may have been as low as 0.4 times the present day LIS at Earth, and up to 1.4 times the present day value deep in the spiral arms. In the absence of hard numbers for the diffusive characteristics in interstellar space, the difference could be substantially higher, or lower, and time dependent. The solar system rotates around the galactic centre and passes through one of the galactic arms every 70–100 million years, and consequently the local interstellar cosmic ray intensity near the solar system would be expected to vary as suggested in Fig. 18.2-2. Furthermore, the recent production of cosmic rays by a supernova may have created localized regions of much higher cosmic ray intensity in the spiral arms that may have been encountered.

The incidence of cosmic rays on a meteorite results in spallation processes that yield both stable and unstable nuclides (Sect. 11.3). The concentration of the stable nuclides increases steadily with time, while an equilibrium concentration is attained for the radioactive nuclides. This allows the exposure age of the meteorite to

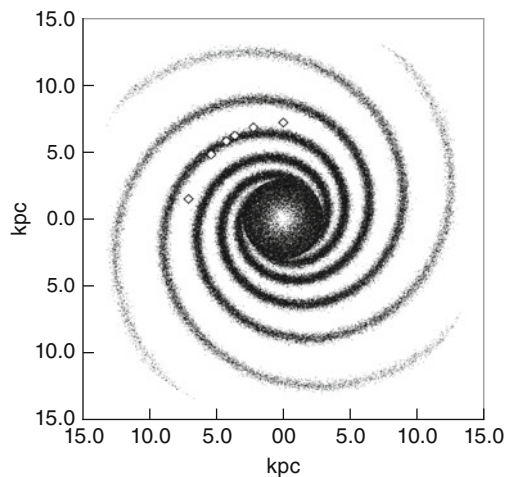


Fig. 18.2-1 A model of our galaxy used to estimate the spatial inhomogeneity of the cosmic ray intensity. Each dot represents a supernova. kpc is the abbreviation for kiloparsec (3,259 light years). After (Busching and Potgieter 2008)

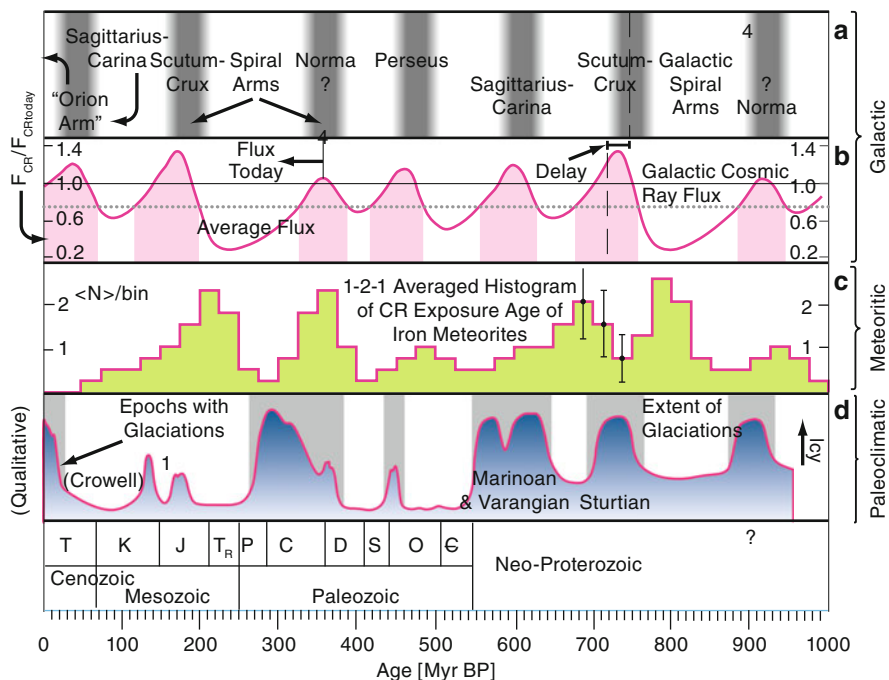


Fig. 18.2-2 The recent cosmic ray history of the solar system (Shaviv 2003). Panel (a): When the solar system passed through the spiral arms of our galaxy; Panel (b): The galactic cosmic ray intensity in the vicinity of the solar system, relative to the present value of the local interstellar spectrum (LIS); Panel (c): The measured exposure ages of Fe/Ni meteorites. Panel (d): Glacial epochs in the geological record

be determined as detailed in Section 23.2.2. If the cosmic ray LIS varies with time, it has been shown that the distribution of exposure ages will cluster about the intervals of lower cosmic ray intensity. Panel (c) of Fig. 18.2-2 plots the observed occurrence of exposure ages of Fe–Ni meteorites and they can be seen to cluster in a manner that is broadly similar to the predicted intervals of lowest LIS in the top panel (Shaviv 2003). Thus the cluster of exposure ages between 170 and 250 Myr might be associated with the low cosmic ray LIS between 200 and 290 Myr; likewise the cluster at 800 Myr with the low cosmic ray intensity in the interval 750–860 Myr. Clearly there are errors in the exposure ages, and the spiral arm crossing times are uncertain. Nevertheless, the broad agreement between panels (a) and (c) shown in Fig. 18.2-2 provides encouragement that considerations such as this may be a profitable way to investigate the environment of the solar system in the distant past.

Note that panel (d) shown in Fig. 18.2-2 presents a qualitative summary of geological information regarding the climate of the Earth in the past. Although highly speculative, this has led some investigators to suggest that there may be a correlation between periods of intense glaciation and the cosmic ray intensity (Shaviv 2003).

To this point we have considered the observable effects of long-term variations in the intensity of the cosmic radiation in the local interstellar environment of the

solar system. Astronomical evidence suggests that the properties of interstellar space have varied substantially with time as well (Mueller et al. 2006). For example, the strength of the local interstellar magnetic field, and the gas density, is expected to vary substantially as the Solar system crosses a spiral arm. Analysis of the absorption lines of the light coming from many stars indicates that interstellar space is populated by a large number of “clouds” with quite strongly varying properties. Frisch and Mueller (2011) estimate that as the solar system passed through various clouds over the past 10^5 years, the ram pressure (the pressure on an object as it moves through a fluid) exerted on the heliosphere may have varied by a factor of 4.2. The evidence suggests that the dimensions of the clouds are such that the Earth will move from one to the next after intervals of the order of 30,000 year and greater.

In Sect. 5.6 we described how the solar wind creates the termination shock when its pressure decreases to that of the local interstellar medium. The relatively low present day interstellar densities and magnetic field strengths resulted in the termination shock (TS) being observed by Voyagers 1 and 2 at 94 and 84 AU, in 2004 and 2007, respectively. If, in the past, the ram pressure of the local interstellar medium were higher, the termination shock would have been closer to the Sun. This, in turn, would reduce the modulating effects on the cosmic rays en route from the TS to Earth. That is, a changing interstellar environment (density, temperature, degree of ionization, or magnetic field) would change the intensity of the galactic cosmic rays at Earth, leading to a change in the production rate of the cosmogenic radionuclides.

The solar system is presently embedded in a warm, low-density interstellar cloud, but astronomical considerations indicate that it moved from a colder, denser cloud ~60,000 years ago and that there may have been a number of similar transitions in the past.

Exploring this possibility, Zank and Frisch (1999) considered the case of the Earth passing through a relatively small (~0.05 parsec) dense molecular cloud. Assuming a 50-fold increase in interstellar gas density, Florinski et al. (2003) estimated that the heliosphere could shrink until the TS was at 22 AU. As a consequence, the modulation of the galactic cosmic rays would decrease, and the intensity could be enhanced by a factor of 1.5 to 4 for energies between 100 MeV and 1 GeV. This would result in an increase in the production rate of all the cosmogenic radionuclides. The model also suggests that the intensity of the anomalous cosmic radiation (ACR) could increase by a factor of ten as a consequence of the higher density of neutral atoms (Box 18.2.1). The predominantly low energy ACR could lead to the preferential production of ^{14}C . At the present, there are no clear examples of either mechanism in the cosmogenic record.

Other ephemeral mechanisms might exist that would result in short-lived (100–1,000 year) increases in the LIS near Earth. Sonett et al. (1987) have considered one such in which a supernova blast wave sweeps past the Solar System, carrying shock accelerated protons and heavier ions that increase the LIS at the solar system as it sweeps by.

In summary, the known large-scale structure of the galaxy, and the exposure ages of meteorites, has led to the inference that the long-term average LIS of the galactic cosmic radiation has varied by a factor of two or more over the past

1,000 Myr, with an approximate periodicity of ~ 150 Myr. In addition, changes in the density and magnetic field strength of the local interstellar medium will have affected the location of the termination shock, leading to changes in the modulation of the cosmic radiation within the heliosphere. There may have been short-lived ($>1,000$ year) increases in the LIS due to other local interstellar effects. All three effects will have led to changes in the production rate of the cosmogenic radionuclides on Earth, and in the solar system. The use of several long, independent, ^{10}Be records may reduce the statistical noise to the point where it may be possible to detect such variations and improve our understanding of the past and present properties of the local interstellar environment.

18.2.1 BOX The Anomalous Cosmic Radiation

Satellite observations by John Simpson and his colleagues in the 1960s showed that the cosmic ray intensities, having decreased steadily with decreasing energy below 1 GeV (Fig. 5.7.3-1), then started to increase rapidly below 50 MeV. Named the “anomalous cosmic radiation (ACR)”, Len Fisk rapidly had the insight to explain this new population of energetic particles.

An anomalous cosmic ray starts its life as a neutral atom of interstellar gas. If it is moving towards the Sun, it can pass through the heliosheath, the termination shock, and the Parker spiral field (Fig. 5.5-4) without being affected by the heliospheric magnetic fields. The intensity of the light from the Sun increases rapidly as the atom approaches the Sun, ultimately leading to photo-ionization into an electron and a positive ion. Being charged, these are deflected by the heliospheric magnetic fields, and start to be swept outwards by the solar wind. Over time the ions are accelerated up to energies of 50 MeV; in a sense this acceleration is the reverse of the process whereby the galactic cosmic rays are decelerated as they move inwards towards the Sun. Clearly, the production rate is determined by the flux of neutral atoms entering the solar system; astronomical data suggest that this may have been a factor of 100 higher on occasions as the Sun orbited the centre of the galaxy (Sect. 18.2). Sometimes the interstellar gas has been highly ionized and then there are few neutral atoms. As a consequence, the intensity of the ACR may have varied over geological time from 100 times the present day intensities, to close to zero.

As discussed in Sect. 10.3.2, production of the majority of the cosmogenic radionuclides requires protons and neutrons with energies greater than the nuclear binding energy (>8 MeV). Ninety-nine per cent of the ^{14}C , however, is produced by thermal neutrons. As a consequence, the majority of the anomalous cosmic rays have sufficient energy to produce ^{14}C , while the efficiency for ^{10}Be , for example, will be much less. Thus large variations in the intensity of the ACR may be reflected in the production rate of ^{14}C in the past. Regrettably, the short half-life of ^{14}C restricts the use of this form of analysis to the past 50,000 years.

18.3 Individual Supernova

To this point we have considered the integrated effects of all the supernova sources in the galaxy (Sect. 5.4), and how the spiral arm structure will result in relatively slowly changing production of cosmogenic radionuclides on Earth, and in meteorites. We now consider the situation where a supernova in the immediate vicinity of the solar system might result in a substantial short-term variation over and above that due to the large-scale structure of the galaxy. As outlined in Sect. 5.4, the known rate of occurrence of supernova in our galaxy suggests that there is a probability of 0.32 that there has been a supernova within 200 parsec within the past 100,000 years. Stated differently, this suggests that ~ 10 supernovae have occurred within 300 parsec in the past million years. It therefore appears quite reasonable to consider whether an SN would have left its mark in the cosmogenic radionuclide record. Table 18.3-1 lists the historical supernovae in the vicinity of the solar system over the past 2,000 years.

We must distinguish two possible effects. The motion of the charged cosmic rays is severely inhibited by the interstellar magnetic field and they will only get to the solar system quickly if it is close to a line of force of the interstellar magnetic field that leads back to the supernova (Sect. 5.4). Otherwise, the cosmic rays must diffuse across the lines of force, and the rise and fall times will be a factor of 10–100 times greater than the time taken for light from the SN to reach Earth. For example; the electromagnetic energy from the Crab at 2.2 kpc ($\sim 6,000$ light years) took $\sim 6,000$ years to reach Earth, so a pulse of charged particles from the Crab SN would have a rise time of $> 60,000$ years.

On the other hand, electromagnetic radiation (e.g., X-rays, ultra-violet, and gamma rays) and neutrons are not deflected by the interstellar magnetic field. They would travel in straight lines to Earth. Further, unlike charged cosmic rays, they would not be deflected by the geomagnetic field. Thus, a pulse of gamma rays from a high declination (north or south) supernova would generate cosmogenic

Table 18.3-1 Historical Supernovae (SN)

Name	Date	Constellation	Distance (kpc)	Visibility (months)	Declination
SN185	185 AD	Centaurus	2.5		62°S
SN386	386 AD	Sagittarius	>5		
SN393	393 AD	Scorpius	~ 10		
SN1006	April, 1006	Lupus	2.2		42°S
SN1054 (Crab)	July, 1054	Taurus	2.2	~ 20	22°N
SN1181	August, 1181	Cassiopeia	8	6	64°N
Vela Junior	1320 $\pm$ 150		~ 0.2		47°S
SN1572 (Tycho)	Nov. 1572	Cassiopeia	2.3	16	64°N
SN1604 (Kepler)	Nov. 1604	Ophiuchus	6.1	12	21°S
Cassiopeia A	about 1700	Cassiopeia	3.4		59°N

These estimates have come from a number of sources. It should be noted that some of the distances in the fourth column vary from one source to another

radionuclides only in northern (or southern) latitudes. As we have seen, there is relatively little transport of the atmosphere between the northern and southern hemispheres (Fig. 13.4-4), so this would result in a considerable difference between cosmogenic responses in the two hemispheres. More than 50% of the galaxy is at declinations $>30^\circ\text{N}$ or S, so there is a good probability that such an N-S difference would be observed. For an SN at a high declination, the existence of an N-S difference in ^{10}Be would be an important test of whether an enhancement in the cosmogenic record was associated with the SN.

The neutron has a half-life of 10.2 min; however, it is known that cosmic rays of $>10^8$ GeV are produced in supernova (Berezinsky et al. 1990). The relativistic time dilation (Box 5.3.1) means that the half-life of a 10^7 GeV neutron (as seen from Earth) is 220 years. Allowing three half lives (attenuation of one-eighth) this suggests some 10^7 GeV neutrons would reach distances >200 parsec from the supernova. However, present day estimates suggest that the production of 10^7 GeV neutrons would be too low to be of significance.

The probability of a detectable gamma ray effect is much larger. It is well known that (p,p) reactions of the cosmic rays in the SN shock front will generate π^0 mesons, which decay immediately to yield two gamma rays ($E > 68$ MeV). Ground level arrays of detectors, and satellites such as COS-B, have detected gamma ray emissions from the Crab pulsar, at a distance of 2.2 kiloparsec. This indicates that the Crab was still a strong source of gamma rays $\sim 1,000$ years after the occurrence of the SN. Theoretical estimates based on the properties of the well-observed supernova SN1989c suggest that the gamma ray pulse from an SN would have had a photon fluence of $\sim 4 \times 10^9$ photons cm^{-2} at a distance of 200 parsec from the SN and that the majority of the emissions would occur within 10^7 s (4 months) (Berezinskii and Ptuskin 1989). The ^{10}Be production cross-sections for high-energy γ -ray-induced spallation of ^{14}N and ^{16}O have been reported to be ~ 10 microbarns (each reaction), indicating that each incident gamma ray has a probability of 8×10^{-5} of initiating photospallation that yields ^{10}Be . The estimated fluence, times the probability, suggests a ^{10}Be production rate of 3×10^5 cm^{-2} . That is comparable to the present day annual production rate. While lacking any accuracy, this estimate suggests that such an SN at 200 pc may have resulted in a significant, short-lived increase in the annual ^{10}Be deposition in one, or both of the polar caps.

Figure 18.3-1 displays a set of data that may possibly be due to a gamma ray burst from such a supernova. The data are the annual ^{10}Be data from Dye 3 (65°N) and North GRIP (75°N) in Greenland, and ~ 6 -year samples from the South Pole. The enhancements near 1459 in the Dye 3 and South Pole records are in excess of five standard deviations in amplitude above the pre-enhancement value: the probability that they are independent local effects that have both occurred in ~ 1459 by chance is very small. The overall setting of the events is displayed in Fig. 7.3-1 and it can be seen that the impulsive enhancements at Dye 3 and South Pole exceed the ^{10}Be values estimated to correspond to the local interstellar spectrum (i.e. there was very little modulation during the Spoerer Minimum). That is, the data suggest that the high values are not explicable in terms of the galactic cosmic radiation. There are two possibilities for an additional source of cosmic rays: (a) a very intense solar flare, (b) a

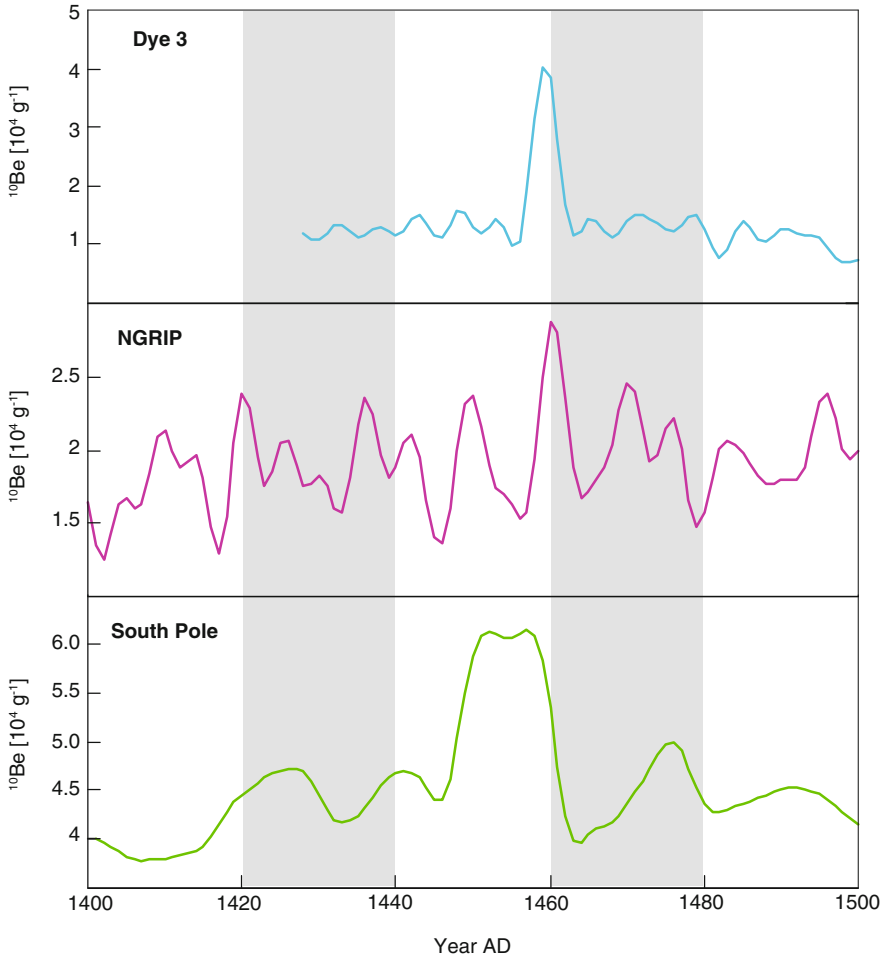


Fig. 18.3-1 The ^{10}Be records from the Arctic and Antarctic in 1400–1500 AD. The enhancements for Dye 3 (southern Greenland) and South Pole represent an ephemeral source of radiation over and above the galactic cosmic radiation. They may be due to a gamma ray pulse from the nearby supernova Vela Junior, or due to a once in a millennium high intensity burst of cosmic rays from the Sun

supernova. Bearing in mind that the two enhanced values at South Pole are 6-year measurements, it can be shown that the total ^{10}Be in the pulse at South Pole was a factor of three times that at Dye 3. While there was an enhancement at NGRIP that was coincident with the Dye 3 event, the high degree of variability of this record means that this only provides minor support to the evidence from Dye 3 and South Pole.

Satellite observations have identified an SN remnant (RXJ0852.0-46.22/GROJ0852-4642), estimated to be at a distance of 200 pc from the solar system, and at a declination of 46.5°S (Aschenbach et al. 1999). It is sometimes called “Vela Junior”. Gamma ray and other observations suggest that it occurred about

1320 AD but with a ± 150 year error box; no visual sighting is known, possibly because it was at a high southern declination. It is significant that it was a tenth of the distance to the closest of the other nine SN that have occurred in the past 2,000 years (Table 18.3-1), so for a given intrinsic brightness it would have been 100 or more times brighter in both light and gamma rays than any of the other historic SN. If any SN were to be sufficiently intense in gamma rays to be seen in the cosmogenic record, this would be it. As noted above, the increase in ^{10}Be production in the Southern hemisphere was three times that in the northern, which is consistent with the declination of 46.5° of the Vela Junior SN remnant.

It must be noted, however, that a solar origin of the enhancement is not impossible; 1459 was in the middle of the Spoerer Minimum of solar activity (Sect. 17.3.1), yet we have seen in Sect. 8.3.3 that there is evidence that large solar energetic particle (SEP) events can occur during periods of low solar activity. Present day experience indicates that large SEP events are usually accompanied by other large SEP events, and that there would have been very strong auroral activity. None such is recorded in the historic records.

In summary, the evidence suggests that the ^{10}Be enhancements in 1459 may be as a result of a gamma ray pulse from Vela Junior. Further observations from both hemispheres would remove the doubt in the future.

Should the evidence indicate that the 1459 event is indeed due to a gamma ray burst from Vela Junior, it indicates that it is worthwhile looking for other weaker SN effects in the cosmogenic data. In particular, it will be desirable to combine several separate records (e.g. using the principal components methodology) to look for SN events that are too small to be identified with certainty in a single record.

References

- Aschenbach B, Iyudin AF, Schonfelder V (1999) Constraints of age, distance and progenitor of the supernova remnant RX J0852.0-4622/GRO J0852-4642. *Astron Astrophys* 350(3):997–1006
- Berezinskii VS, Ptuskin VS (1989) Diffusive shock acceleration of cosmic ray and gamma radiation from young supernova shells. *Astrophys J* 340:351–354
- Berezinsky VS, Bulanov SV, Dogiel VA, Ginzburg VL, Ptuskin VS (1990) *Astrophysics of cosmic rays*. North Holland, Amsterdam
- Busching I, Potgieter MS (2008) The variability of the proton cosmic ray flux on the Sun's way around the galactic center. *Adv Space Res* 42(3):504–509
- Florinski V, Zank GP, Axford WI (2003) The Solar System in a dense interstellar cloud: implications for cosmic-ray fluxes at Earth and Be-10 records. *Geophys Res Lett* 30:GL2206
- Frisch PC, Mueller H-R (2011) Time variability in the interstellar boundary conditions of the Heliosphere over the past 60,000 years: impact of the solar journey on the galactic cosmic ray flux at Earth. *Space Sci Rev* (in press)
- Mueller HR, Frisch PC, Florinski V, Zank GP (2006) Heliospheric response to different possible interstellar environments. *Astrophys J* 647(2):1491
- Rood RT, Sarazin CL, Zeller EJ, Parker BC (1979) X-rays or gamma-rays from super-novae in glacial ice. *Nature* 282(5740):701–703

- Shaviv NJ (2003) The spiral structure of the Milky Way, cosmic rays, and ice age epochs on Earth. *New Astron* 8(1):39–77
- Sonett CP, Morfill GE, Jokipii JR (1987) Interstellar shock-waves and Be-10 from ice cores. *Nature* 330(6147):458–460
- Zank GP, Frisch PC (1999) The interaction of the solar wind with a high density cloud. *AIP Conf Proc* 471:831–834

Chapter 19

Atmosphere

19.1 Introduction

As discussed in Chap. 13, the atmosphere consists of several distinct layers. The bottom layer, the troposphere, is where we live and where the weather takes place. It is a very dynamic system which reaches up to about 16–18 km at low latitudes and 8–12 km in the polar regions. Large amounts of energy and water are transported in the troposphere by convection and winds. As its name indicates the second layer, the stratosphere, is stratified and less dynamic. It ranges up to 40 km and together with the troposphere it accounts for more than 90% of the total atmospheric mass. The stratosphere and troposphere are separated by the tropopause. On the average it takes 1–2 years for a parcel of stratospheric air to descend into the troposphere. The exchange between stratosphere and troposphere is complex and shows seasonal fluctuations which are poorly understood and difficult to model. The residence time of aerosols in the troposphere is comparatively short (1–3 weeks).

Radionuclides have considerable potential for improving our understanding of atmospheric dynamics. There are two main reasons:

1. The characteristics of several sources of radionuclides are well known and the relationship between them, and the observed fluxes of radionuclides throughout the world, can be used to quantify atmospheric dynamics and to validate transport models. One important and special feature is that there are three distinctly different sources of radionuclides in the atmosphere, with quite different temporal and spatial characteristics. The first, the cosmogenic radionuclides, have a well-defined source function that is distributed in a non-homogeneous manner over the whole atmosphere with the greatest production in the stratosphere at high latitudes (Chap. 10). The second source consists of a number of short-lived injections at known locations such as nuclear bomb tests and accidental releases from nuclear power plants (e.g. Chernobyl, Fukushima). The third class consists of the continuous emissions from nuclear reprocessing plants (La Hague, Sellafield). In the following we describe atmospheric studies based on the first two of these sources.

2. Radionuclides with half-lives similar to the time scales of atmospheric processes (e.g. ^7Be , $T_{1/2} = 53.2$ days) offer an additional opportunity to study the dynamics of atmospheric processes. This is the “clock” property which is unique for radioactive tracers.

19.2 Studies of Atmospheric Mixing

As a result of geomagnetic shielding, the cosmic ray flux is largest at high latitudes (Sect. 5.8). This is well demonstrated by the latitude dependence of the cosmogenic radionuclide production rate which starts at a low level between the equator and about 30° , increases rapidly up to $\sim 50^\circ$, and reaches its maximum at about 70° from where it stays constant (Fig. 10.3.3-1). An initial estimate of the characteristics of global mixing can therefore be obtained by measuring the concentration of cosmogenic radionuclides in the lower troposphere (e.g. ^{14}C) and from the latitudinal distribution of the deposition fluxes (e.g. ^{10}Be). ^{14}C measurements in tree rings reveal almost worldwide uniformity, indicating almost complete global mixing. There is a small depletion in the southern hemisphere of about 2–4 permil which is not the result of an incomplete atmospheric mixing, but rather as a result of more effective exchange of CO_2 with the ocean which covers about 70% of the southern hemisphere compared with 50% in the northern hemisphere. The mean global atmospheric residence time of ^{14}C with respect to gas exchange with the ocean is 7–8 years. It can therefore be concluded that on time scales longer than 10 years the troposphere and stratosphere are well mixed.

Considering cosmogenic radionuclides which attach to aerosols and are removed from the atmosphere mainly by wet precipitation (^{10}Be , ^{26}Al , ^{36}Cl) the situation is quite different. First, the residence times are much shorter ranging from a few years in the stratosphere to a few days to several weeks in the troposphere. Second, the removal is strongly governed by the precipitation rate. Third, because of the short tropospheric residence time, the pathway of the aerosol bound cosmogenic radionuclides from the stratosphere into the troposphere plays an important role.

Again, we start with some simple general considerations. As already discussed elsewhere (Chap. 13), without atmospheric mixing, changes in the strength of the geomagnetic dipole field would only affect the precipitation rate of the cosmogenic radionuclides at low latitudes ($<60^\circ$). Production would not be affected at high latitudes, and the cosmogenic radionuclide concentration in polar ice would not change, in clear contrast to the rather large variations that are observed (Sect. 21.2). On the other hand, the effect of solar modulation (11-y Schwabe solar cycle, Grand Minima) would be greater at high latitudes where the low-energy cosmic rays (which are particularly sensitive to solar modulation) can enter the atmosphere, while the geomagnetic field prevents the from reaching the equatorial regions (see Stoermer cut-off rigidity, Eq. 5.8.2-2). The ^{10}Be observations make it clear that there is a degree of mixing within the 1–3 weeks before the aerosols carrying the

cosmogenic nuclides are scavenged from the atmosphere. This was discussed in greater detail in Chap. 13.

So far we have only discussed radionuclides (^{14}C , ^{10}Be) with half-lives which are much longer than the typical time constants of atmospheric dynamics. We now consider ^7Be , whose half-life of 53.2 days makes it well suited to the study of atmospheric transport and mixing processes. A difficulty in interpreting a single cosmogenic radionuclide such as ^7Be is that the atmospheric production rate is strongly depth- and latitude-dependent. This problem can be avoided to a large extent by using the $^{10}\text{Be}/^7\text{Be}$ ratio, which is much more constant with latitude than the production rate (Fig. 10.3.3-5).

To demonstrate the potential of the $^{10}\text{Be}/^7\text{Be}$ ratio we use a simple 2-box model consisting of the stratosphere and the troposphere (Fig. 19.2-1). To keep things simple we assume a constant $^{10}\text{Be}/^7\text{Be}$ production ratio of 0.5. In reality, when averaged over all latitudes, it ranges from 0.4 at the top of the atmosphere to about 0.9 at sea level. The average atmospheric ratio changes from 0.65 at low latitudes to 0.71 at high latitudes for low solar activity (Sect. 23.2.3). Unlike the box diffusion model we used in the case of ^{14}C (Fig. 13.5.3.1-1), this model includes a steady deposition rate of $1/(0.058 \text{ year})$ that quantifies the 3 week ($= 0.058 \text{ year}$) residence time of aerosols in the troposphere.

In our first case study we simulate the 11-y Schwabe cycle with a sinusoidal production function for ^7Be with a peak-to-peak amplitude of 30% and a mean production rate of 100. Figure 19.2-2 shows the results. Panel (a) depicts the total ^7Be production in the stratosphere and the troposphere for a period of 25 years. The corresponding ^{10}Be production is half as large (not shown) and the ^{10}Be deposition flux from the atmosphere to the Earth's surface (panel b) is, as expected, equal to the production rate but delayed by about 1 year because of the stratospheric residence time of 2 years and the tropospheric residence time of ~ 3 weeks. The ^7Be deposition flux (panel c) is smaller than the production because practically all

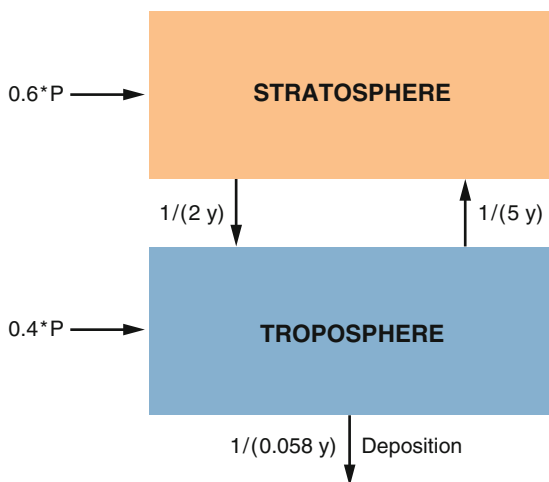


Fig. 19.2-1 Simple 2-box model to demonstrate the potential use of the $^{10}\text{Be}/^7\text{Be}$ ratio as a tracer of atmospheric mixing

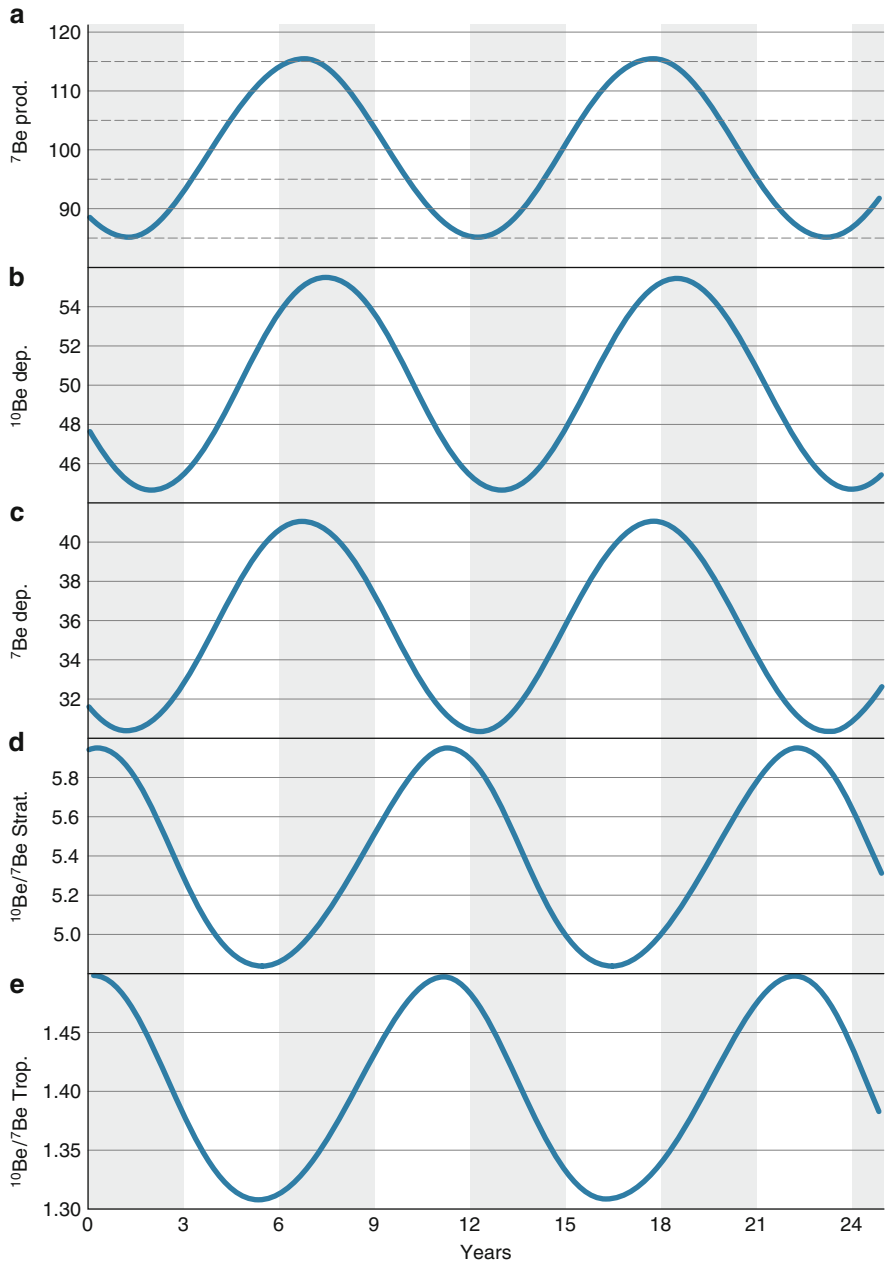


Fig. 19.2-2 Simulation of the $^{10}\text{Be}/^7\text{Be}$ ratio with the 2-box model of Fig. 19.2-1 for a period of 25 years. (a) Total ^7Be production. The total ^{10}Be production (not shown) is half as large; (b) ^{10}Be deposition; (c) ^7Be deposition; (d) $^{10}\text{Be}/^7\text{Be}$ in the stratosphere; (e) $^{10}\text{Be}/^7\text{Be}$ in the troposphere

of the stratospheric production (60% of the total) decays before reaching the troposphere. A small part of the 40% of the ^7Be produced in the troposphere decays, leaving, on average 36% of the ^7Be production which is then scavenged from the atmosphere. The $^{10}\text{Be}/^7\text{Be}$ ratio of 5.4 in the stratosphere (panel d) is about a factor of ~ 11 larger than the production ratio of 0.5 as a consequence of the decay of the ^7Be . It fluctuates with a peak-to-peak amplitude of $\sim 22\%$ and is delayed by about 5 years compared to the ^7Be production. Finally the $^{10}\text{Be}/^7\text{Be}$ ratio in the troposphere (panel e) has a mean value of about 1.5 and an amplitude of 14%. These results show that stratospheric air is characterized by a significantly higher $^{10}\text{Be}/^7\text{Be}$ ratio than that from the troposphere (5.4 compared with 1.5) and that the intrusion of stratospheric air parcels into the troposphere should be detectable based on their higher $^{10}\text{Be}/^7\text{Be}$ ratio.

In a second case study we simulate the effect of a variable stratosphere–troposphere exchange (STE) for a period of 25 months (Fig. 19.2-3).

The production rate is kept constant at 100 units for ^7Be and 50 for ^{10}Be . As before, we assume that 60% of the production is in the stratosphere, and 40% in the troposphere. The exchange processes between the stratosphere and the troposphere (quantified by an exchange parameter k_{ST} which is the inverse of the residence time) are assumed to show a seasonal sinusoidal change with a mean value of 0.5 and a peak-to-peak amplitude of 0.5 (panel a) with a higher exchange rate in summer than in winter. In the stratosphere the ^{10}Be fluctuates around 61 units with an amplitude of about 7.4% (panel b). The fact that it is slightly higher than the mean production of 60 units for the stratosphere is as a result of a flux from the troposphere feeding back into the stratosphere. The peak-to-peak amplitude is $\sim 7.4\%$ with a lag of 8.5 months relative to k_{ST} . The mean value of ^7Be in the stratosphere is 11.5 units (compared to production of 60 units) and the peak-to-peak amplitude is 5.7% and the lag 7.5 months (panel c). In the stratosphere the $^{10}\text{Be}/^7\text{Be}$ ratio has a mean value of 5.33, an amplitude of 4.7% and a lag of 9.5 months (panel d). The last panel (e) shows that the $^{10}\text{Be}/^7\text{Be}$ ratio in the troposphere varies sinusoidally around a mean of 1.4 with a peak-to-peak amplitude of 50% and almost in phase with k_{ST} .

These case studies show that ^7Be , taken together with ^{10}Be as a control, has great potential in the study of atmospheric transport and in particular in elucidating the exchange processes between stratosphere and troposphere.

Figure 19.2-4 presents the $^{10}\text{Be}/^7\text{Be}$ ratios measured in monthly rain samples obtained in Switzerland (Heikkilä et al. 2008). The first collection site is the high alpine research station Jungfraujoch (3,200 m asl) while the other station is situated next to the Swiss Federal Institute of Aquatic Science and Technology (Eawag) in Dübendorf (440 m asl) near Zurich. The monthly $^{10}\text{Be}/^7\text{Be}$ ratios for the period 1998–2005 show similar ratios, ranging with a few exceptions from 1.5 to 4. This range agrees nicely with the model results showing mean global tropospheric ratios above 1 and stratospheric ratios below 5.5 (Fig. 19.2-3). There is a clear seasonal trend pointing to an increased air exchange between stratosphere and troposphere during spring and early summer.

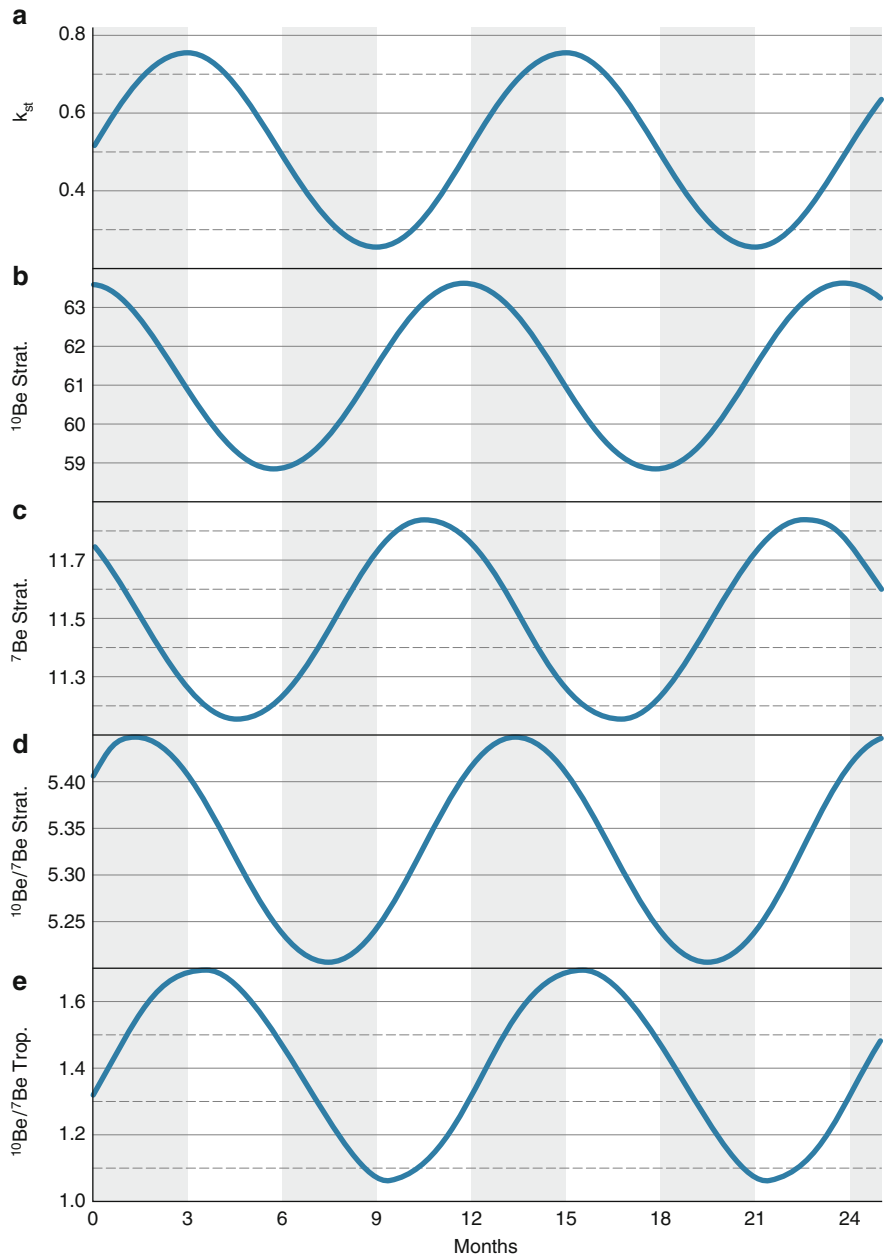


Fig. 19.2-3 Simulation of a variable stratosphere–troposphere exchange (STE) for a period of 25 months. (a) Stratosphere–troposphere exchange coefficient; (b) stratospheric ^{10}Be concentration; (c) stratospheric ^7Be concentration; (d) stratospheric $^{10}\text{Be}/^7\text{Be}$ ratio; (e) tropospheric $^{10}\text{Be}/^7\text{Be}$ ratio

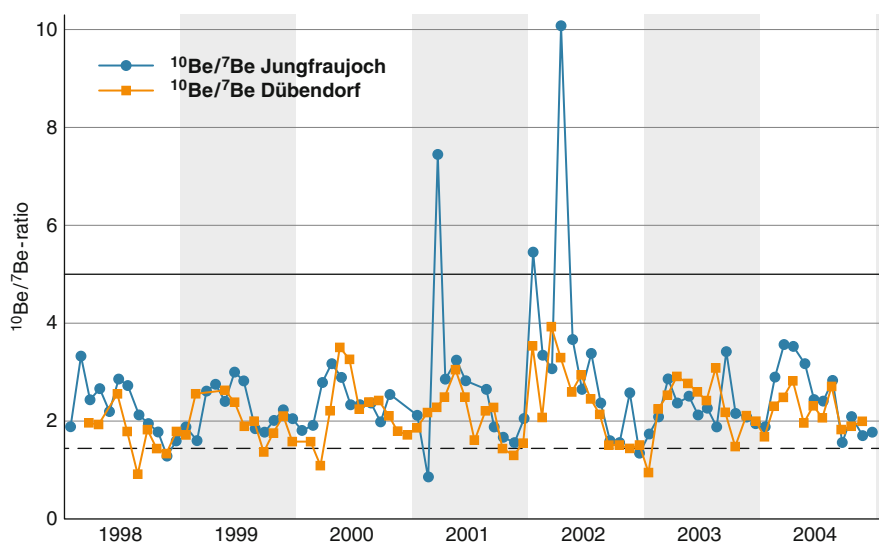


Fig. 19.2-4 Results of $^{10}\text{Be}/^{7}\text{Be}$ ratio measurements in the precipitation in Switzerland at 440 m (Dübendorf) and 3,200 m (Jungfraujoch) above sea level. After (Heikkilä et al. 2008)

A better understanding of the STE is crucial in atmospheric research because it strongly affects chemical concentrations such as ozone and processes related to small- and large-scale air exchange. The large difference in the $^{10}\text{Be}/^{7}\text{Be}$ ratio between stratosphere (~ 6) and troposphere (~ 1) makes this ratio an excellent tracer to study these processes and to improve the present models.

19.3 ^{36}Cl Bomb Pulse as a Tracer of Atmospheric Transport

Following the development of nuclear bombs in the 1940s a large number of tests were performed by a growing number of nations (Fig. 19.3-1). In the first phase these tests were mainly carried out in the atmosphere or in the oceans and they introduced a large quantity of radioactive material into the environment. To stop this process an international treaty banning atmospheric tests was developed and signed by most of the involved nations on August 5, 1963. The last atmospheric tests took place on November 4, 1962. From 1963 onwards the number of tests continued to grow but because, with a few exceptions, they took place underground the release of radionuclides into the atmosphere was significantly reduced. In October 1958 the United States announced a unilateral testing moratorium which was also respected by the former Soviet Union. The tests were resumed in September 1961.

One of the bomb-produced radionuclides, ^{36}Cl , is a useful tracer for studying both atmospheric and hydrological processes (Sect. 20.5). ^{36}Cl is continuously

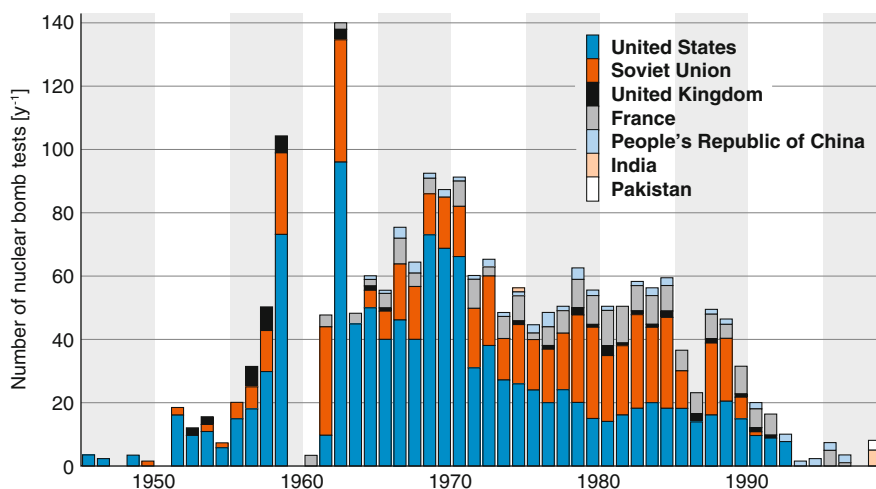


Fig. 19.3-1 The number of annual nuclear bomb tests between 1945 and 1998, colour-coded according to the countries which carried them out. After 1963, with a few exceptions all tests were conducted underground releasing only gaseous radionuclides to the atmosphere (Wikimedia Commons)

produced in the atmosphere by the interaction of cosmic rays with argon; however, the mean global natural production rate is small because of the low abundance of argon (mean global production rate = $1.12 \times 10^{-3} \text{ cm}^{-2} \text{ s}^{-1}$). During the bomb tests the production of ^{36}Cl increased by almost three orders of magnitude, which provides a relatively sharp and very strong signal that was then sequestered in ice, worldwide. It may be surprising that the peak of the ^{36}Cl bomb pulse precedes the peaks of the ^{137}Cs and other nuclear fission products in these archives. The reason is that ^{36}Cl is produced by neutron activation of ^{35}Cl and the production of ^{36}Cl therefore depends strongly on the presence of ^{35}Cl at the test site. Many of the earlier tests were carried out on atolls and ships close to or even in seawater with high chlorine content. Later, the tests were moved inland and restricted mostly to underground sites (Nevada, Utah, Novi Sibirsk, Novaja Semlya) and the production of ^{36}Cl was greatly reduced.

There are eight records of ^{36}Cl covering the period 1950–1975 with high temporal resolution (mostly annual), six from the Northern Hemisphere and two from the Southern Hemisphere. The drill sites are shown on the map (Fig. 19.3-2).

Two cores were from Greenland (Dye 3 and NGRIP) (Elmore et al. 1982; Synal et al. 1990), two from the Swiss alps (Grenzgletscher and Fiescherhorn) (Eichler et al. 2000), two from Central Asia (Tien Shan and Guliya) (Green et al. 2004; Thompson et al. 1995), and one each from the Andes (Huascaran) and Antarctica (Berkner island). On the basis of the measured ^{36}Cl concentrations, the dating of the individual cores and the estimated accumulation rates, the corresponding ^{36}Cl fluxes were calculated (Fig. 19.3-3 upper panel). All the records reach their maxima shortly before 1960 and show fluxes which differ by about a factor of 20. Heikkilä

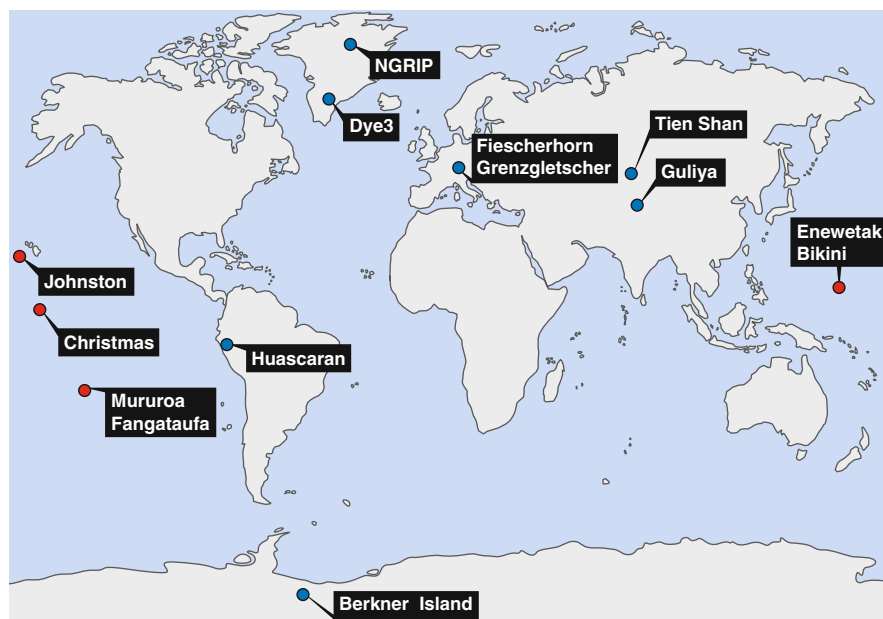


Fig. 19.3-2 Map showing the weapon test sites (red dots) that were near the sea, and the drill sites (blue dots) of the ice cores which were analysed for ^{36}Cl

et al. (2009) modelled the observed fluxes (Fig. 19.3-3, lower panel) by applying the general circulation model ECHAM5-HAM (Sect. 13. 4). A total of 31 documented nuclear explosions that were near the sea were used for this study.

Unfortunately no precise information is available about the amount of ^{36}Cl injected into the stratosphere. It has therefore been estimated based on the magnitude (energy output) of each bomb test, yielding estimates of between 1 and 45 kg of ^{36}Cl per bomb. The modelled ^{36}Cl fluxes are in good general agreement with the observations. Thus the ~20-fold difference observed between the Antarctic and northern Europe and the Arctic is reproduced by the model, providing confidence that the model treats the distribution and scavenging of ^{36}Cl properly (and also showing that the estimated amounts of ^{36}Cl are reasonable). Some features, such as the slower rise to maximum of the observations and the dependence on latitude in the Northern Hemisphere may provide further information on atmospheric transport in the future.

The upper panel of Fig. 19.3-3 shows the measured ^{36}Cl fluxes in ice cores from Greenland (NGRIP, Dye 3), the Alps (Fiescherhorn and Grenzgletscher), the Himalayas (Guliya, Tien Shan), the Andes (Huascarán), and Berkner Island in Antarctica. The lower panel presents the corresponding results of the model runs using the estimates of the injected ^{36}Cl as indicated by the vertical lines on the time axis. The largest deposition fluxes are expected and observed at lower latitudes

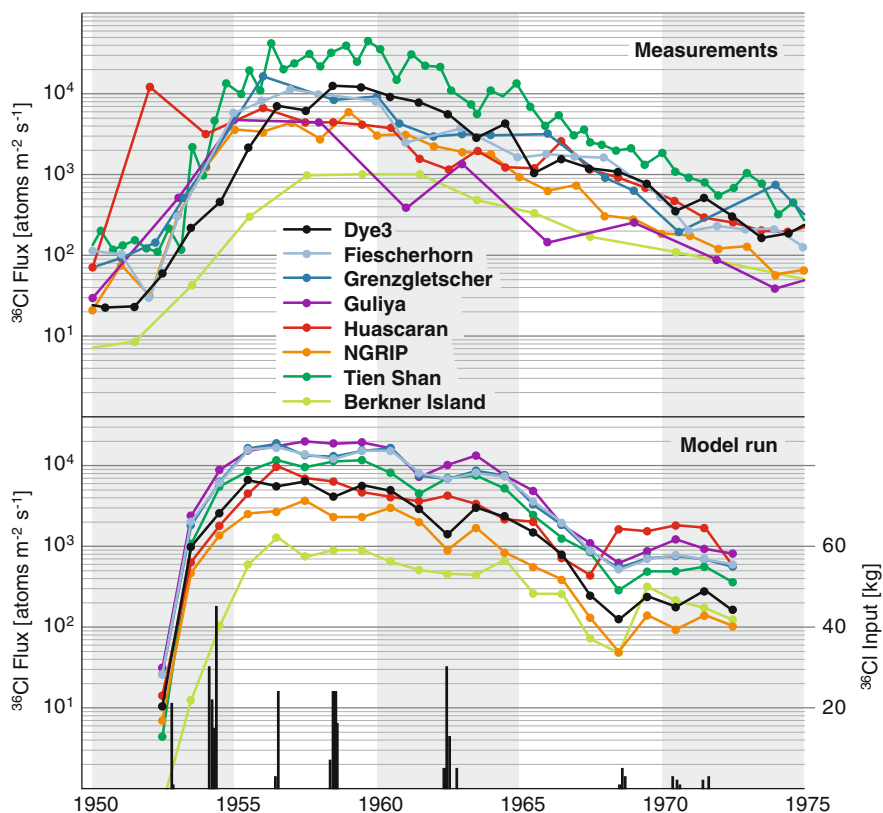


Fig 19.3-3 Comparison of measured (*upper panel*) and modelled (*lower panel*) ^{36}Cl fluxes for North Greenland (NGRIP), South Greenland (Dye 3), the Alps (Fiescherhorn and Grenzgletscher), the Himalayas (Gulya, Tien Shan), the Andes (Huascarán), and Berkner Island in Antarctica. The vertical black lines in the *lower panel* indicate the time and estimated yields of the nuclear test explosions

while the smallest are found in remote polar regions. The ratio between the two is more than 20:1.

The ^{36}Cl peak activity was reached shortly before 1960. The decreasing trend afterwards is clearly interrupted by additional ^{36}Cl input from later test explosions. Even the relatively weak inputs around 1970 are visible. From the slope between the tests the atmospheric residence time can be estimated.

19.4 Concentrations and Fluxes

As discussed in previous chapters, the time series of cosmogenic radionuclide measurements obtained from a natural archive reflects the combined effects of production and system changes. The production changes are because of the

modulation of the primary galactic cosmic rays by solar activity and the geomagnetic field intensity (Sects. 5.7 and 5.8). The production signal, which shows a specific altitude and latitude dependence, is modified considerably by atmospheric transport and deposition processes (Chaps. 10 and 13). This raises two questions:

1. How can we distinguish between production and system contributions to the observed variations? This question is of fundamental importance when using cosmogenic data for many purposes and is discussed in the Chaps. 17–22.
2. A more specific question is how to relate the measured concentration of ^{10}Be , ^{36}Cl , or ^{26}Al in a natural archive to the atmospheric production rate in the past.

In practice, analysis of the cosmogenic radionuclides in a sample always commences with a measurement of the concentration, i.e. the number of atoms per gram of the sample material (Chap. 15). This was determined by the number of atoms deposited at the sampling site in a year, for example, and the amount of precipitation (snow, ice) in which these atoms are stored in. Consequently some (but not all) changes in the precipitation rate lead to ^{10}Be concentration changes which have nothing to do with ^{10}Be production changes. In these cases it may be more appropriate to use the ^{10}Be flux instead of the ^{10}Be concentration. As an example the annual ^{10}Be flux in an ice core is given by

$$F = \rho ac \quad (19.4-1)$$

with F the ^{10}Be flux in $\text{atoms cm}^{-2} \text{ year}^{-1}$, ρ the density of ice (0.92 g cm^{-3}), a the ice accumulation rate in cm year^{-1} , and c the ^{10}Be concentration in atoms g^{-1} .

Clearly, the flux F tells us the rate at which ^{10}Be atoms were deposited per unit area and time. At first sight it may appear that the flux is *a priori* independent of the precipitation rate (ρa). We now use two extreme situations to demonstrate that neither the concentrations nor the fluxes are perfect measures of the production rate, and that it depends on the specific conditions whether concentration or flux is preferable. The fact that neither concentration nor flux reflects the production rate correctly is one of the limiting factors of the cosmogenic radionuclide method. However, there are ways to circumvent this problem at least in part, and we discuss them at the end of this chapter. Furthermore, the production changes are frequently large enough (e.g. the Grand Minimum response, Fig. 7.3-1) that the investigations of interest can be made using either concentration or flux.

Consider first the situation where the mean global precipitation rate (of snow and rain) increases, while the ^{10}Be production rate remains constant. This results in dilution of the ^{10}Be in an ice sample, and the ^{10}Be concentration therefore decreases. Clearly, in this case, the decrease in concentration is because of a system effect. Figure 19.4-1 illustrates such a situation. In the vicinity of $\sim 12,000$ BP the measured ^{10}Be concentration drops by a factor of two from a high value ($\sim 40 \times 10^3 \text{ atoms g}^{-1}$) to $\sim 20 \times 10^3 \text{ atoms g}^{-1}$ (Finkel and Nishiizumi 1997). This decrease occurred in synchronism with the transition from the last glacial epoch to the Holocene expressed by $\delta^{18}\text{O}$ (Fig. 19.4-1). The good correlation between ^{10}Be concentration and $\delta^{18}\text{O}$ clearly points to a causal relationship. In fact, it is known

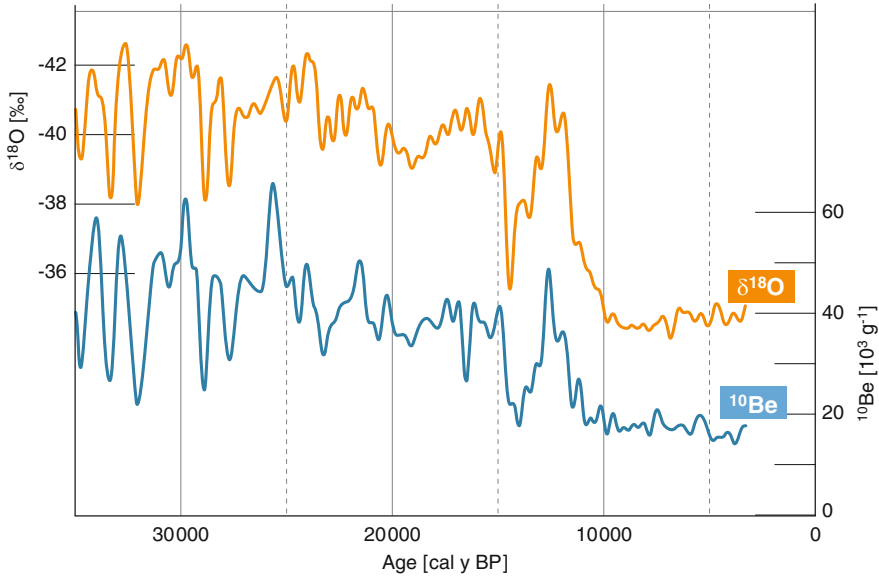


Fig. 19.4-1 Comparison between the ^{10}Be concentration and $\delta^{18}\text{O}$ in the GISP 2 ice core (Finkel and Nishiizumi 1997) for the period 3000–35000 years BP. Note the jump in ^{10}Be concentration from a relative constant value of $20 \times 10^3 \text{ atoms g}^{-1}$ during the Holocene to values up to $60 \times 10^3 \text{ atoms g}^{-1}$ during the last glacial when colder and drier conditions prevailed

from other evidence that during cold periods (glacial times) the water cycle was reduced and as a consequence the ice accumulation rate in the polar region was only about 50% of today's value. However, the generally cold glacial climate was interrupted with warm episodes, the so-called Dansgaard-Oeschger events, probably the result of changes in the thermohaline circulation, resulting in abrupt changes in temperature and precipitation rates as well. That is, the high ^{10}Be concentrations during the last glacial period were a direct consequence of the lower precipitation rate at that time. Assuming constant production, we would predict a decrease in ^{10}Be concentration of a factor of two as observed at the commencement of the Holocene.

As a second extreme case, we now discuss the effect of very local system changes on the cosmogenic data. As in the previous case, we assume that the ^{10}Be production rate remains constant. We consider the effect of a small shift in the local precipitation pattern without any change of the mean global value. For example, assume that the clouds that scavenge the radionuclides change their trajectories slightly, and 20% more of the ^{10}Be bearing snowflakes are delivered to the sampling site at the expense of the surrounding regions, which get correspondingly less. In this case, the ^{10}Be concentration in the ice remains unchanged; the ice accumulation has increased by 20%; and from (19.4-1) the computed flux has increased by 20%. As we have explained, however, this increase has nothing to do with production changes, showing that flux is not a perfect measure of the production rate in this case.

The reason why neither concentrations nor fluxes are perfect is because the ^{10}Be deposition and precipitation rates are coupled (see wet deposition in Sect. 13.3.2). If they were not coupled, as in the case of the dry deposition (Sect. 13.3), the flux would be the correct measure of the production rate. If however ^{10}Be is, as usual, contained in the water droplets in the cloud, then more precipitation will inevitably lead to a larger flux. In practice the situation may be even more complex in that the rate at which a cloud loses ^{10}Be may be highest at the start of a precipitation event, and then decrease slightly with time.

So what can be done to reduce the uncertainty introduced by these deposition (system) effects? There are two approaches. If we know that climatic changes have occurred over large scales (large regions, hemisphere, global) then fluxes are the better choice as is the case for the last glacial epoch (Fig. 19.4-1). When the precipitation changes occur on small scales (local, regional), concentrations are generally preferable. In the second case, the concentration and flux records often show a high degree of similarity and both concentration and flux will produce similar results. Independent of the choice there always is some local noise in the data.

A useful strategy to reduce both the regional and local noise is to average the data from different sites or to apply a more sophisticated technique such as principal component analysis (Jolliffe 2002). An even more satisfactory approach is to combine a ^{10}Be ice core record with a ^{14}C tree ring record. As both ^{10}Be and ^{14}C are produced in a similar manner, but experience significantly different transport mechanisms from the atmosphere into the respective archive (Chap. 13), the common production signal can be extracted using principal component analysis or an equivalent mathematical technique. The second approach is to use a general circulation model to simulate the atmospheric transport and deposition of ^{10}Be . In principle this approach takes care of all the processes affecting the measured concentration. However, it requires large amounts of computer time, detailed climate information about the past, which may not be available, and a very fine spatial resolution, all of which makes this approach difficult. Nevertheless, climate model runs can be very useful for case studies (see the model results for the Laschamp event in Sect. 13.4).

References

- Eichler A, Schwikowski M, Gaggeler HW, Furrer V, Synal HA, Beer J, Saurer M, Funk M (2000) Glaciochemical dating of an ice core from upper Grenzgletscher (4200 m a.s.l.). *J Glaciol* 46(154):507–515
- Elmore D, Tubbs LE, Newman D, Ma XZ, Finkel R, Nishiizumi K, Beer J, Oeschger H, Andrée M (1982) ^{36}Cl bomb pulse measured in a shallow ice core from Dye 3, Greenland. *Nature* 300:735–737
- Finkel RC, Nishiizumi K (1997) Beryllium-10 concentrations in the Greenland ice sheet project 2 ice core from 3-40 ka. *J Geophys Res* 102(C12):26699–26706

- Green JR, Cecil LD, Synal HA, Santos J, Kreutz KJ, Wake CP (2004) A high resolution record of chlorine-36 nuclear-weapons-tests fallout from Central Asia. *Nucl Instrum Methods Phys Res B* 223:854–857
- Heikkilä U, Beer J, Alfimov V (2008) Beryllium-10 and beryllium-7 in precipitation in Dubendorf (440 m) and at Jungfraujoch (3580 m), Switzerland (1998–2005). *J Geophys Res-Atmos* 113(D11).
- Heikkilä U, Beer J, Feichter J, Alfimov V, Synal HA, Schotterer U, Eichler A, Schwikowski M, Thompson L (2009) Cl-36 bomb peak: comparison of modeled and measured data. *Atmos Chem Phys* 9(12):4145–4156
- Jolliffe IT (2002) *Principal component analysis*, 2nd edn. Springer, New York
- Synal HA, Beer J, Bonani G, Suter M, Wolfli W (1990) Atmospheric transport of bomb-produced Cl-36. *Nucl Instrum Methods Phys Res B* 52(3–4):483–488
- Thompson LG, Mosley-Thompson E, Davis ME, Lin PN, Dai J, Bolzan JF, Yao T (1995) A 1000 year climate ice-core record from the Guliya ice cap, China: its relationship to global climate variability. *Ann Glaciol* 21:175–181

Chapter 20

Hydrosphere

20.1 Introduction

Water is vital for the existence of life on Earth. It is also one of the most important inputs to, and outputs from, global change. Access to drinking and agricultural water is still one of the fundamental and vexatious problems of mankind. Understanding all the different aspects of water is therefore crucial and a central topic in applied sciences. In this section we discuss some selected examples that illustrate how the cosmogenic radionuclides contribute to the solution of some of the problems and challenges that we face.

Water participates in a diverse number of processes, spread widely in space and stretching over long periods of time that we call the hydrological cycle. It is driven by solar power and climatic factors (Fig. 20.1-1), and it has consequently varied between the glacial and interglacial epochs. Evaporation of seawater results in about $4.34 \times 10^5 \text{ km}^3$ of water entering the atmosphere each year (Table 20.1-1) and about 92% returns rather rapidly to the sea as oceanic rain. Only 8% reaches the continents, where it, together with the portion of water evaporated from the land, precipitates in the form of rain and snow. Part of that precipitation becomes surface runoff and enters the rivers and lakes, and most of it returns back to the sea within a few months. Snow in the mountains usually melts in spring and summer, providing an important source of water in areas with low summer precipitation, and then returns to the sea. Another part of the continental precipitation evaporates from the lakes, soil, and vegetation (evapotranspiration). Finally, a third part, usually called “groundwater”, percolates into the soils and underlying “aquifers” in porous rocks such as sandstone and faulted regions of less porous rocks such as igneous and metamorphic rocks (e.g. granite and dolerite). This water then generally moves very slowly (metres per year) and may remain in the aquifers for millions of years. The groundwater feeds springs, infiltrates into rivers, and ultimately reaches the sea, thereby closing the hydrological cycle. If the total amount of water evaporated were equally distributed over the Earth’s surface the average precipitation rate would be 1 m per year. If one considers only the continents the mean annual precipitation rate is about 0.6 m.

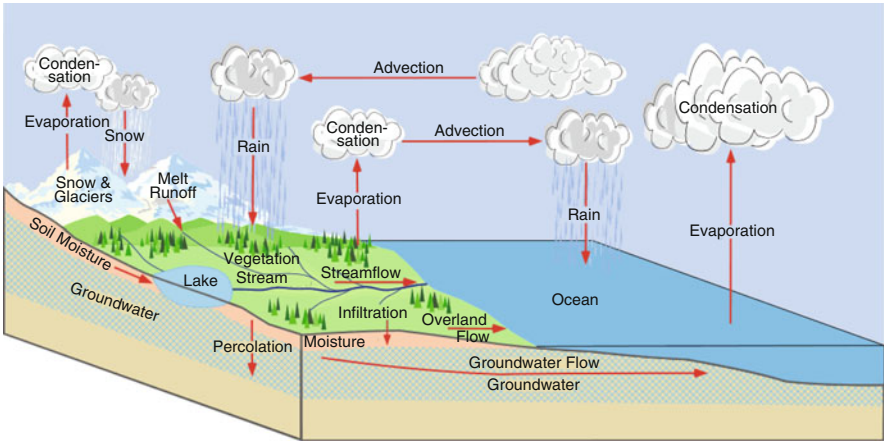


Fig. 20.1-1 The hydrological cycle

Table 20.1-1 Main fluxes of water within the hydrological cycle

Water flux	Annual rate ($10^3 \text{ km}^3 \text{ y}^{-1}$)
Evaporation from oceans	434
Precipitation over oceans	398
Evaporation from land	71
Precipitation over land	107
Runoff and groundwater from land	36

(http://www.newworldencyclopedia.org/entry/Water_cycle)

It is worth noting that it requires about $4 \times 10^{16} \text{ W}$ to evaporate the $500 \times 10^3 \text{ km}^3$ of water per year from the ocean and the land. This corresponds to 30% of the total power received from the Sun, and is 2,000 times greater than the present global power consumption by human society.

Table 20.1-2 lists the major water reservoirs, and their absolute and relative sizes with respect to the global total, and to the freshwater component, respectively. In the last column the range of the residence times is given. It should be noted that several of these numbers (e.g. size of groundwater reservoir) are subject to considerable differences in the literature.

While the total hydrosphere contains about $1.4 \times 10^6 \text{ km}^3$ of water, only about 2.75% of it is freshwater, the rest (97.25%) consisting of saline water. Interestingly most of the freshwater is not stored in lakes and rivers, but frozen in the ice sheets of Antarctica and Greenland and the alpine glaciers (75%) and in the underground aquifers (~25%). Lakes and rivers account for less than 1% of the fresh water.

The residence time of the water in the various reservoirs is calculated by dividing the reservoir size by the total outflow (evaporation, rivers, etc.). However, note that this only represents a mean value. The variability within a reservoir can be very large. For example, evaporation only takes place on the ocean surface and the oceanic rain brings most of the water back to the surface within hours. Therefore the

Table 20.1-2 Sizes of the main water reservoirs and typical residence times

Reservoir	Volume (10 ⁶ km ³)	Percent of total water	Percent of freshwater	Residence time
Oceans	1,370	97.25	0.0	3,200 years
Ice caps and Glaciers	29	2.05	75	20–100/10,000 years
Groundwater	9.5	0.68	24.5	10,000 years
Lakes	0.125	0.01	0.3	50–100 years
Soil moisture	0.065	0.005	0.2	1–2 months
Atmosphere	0.013	0.001	0.03	10 days
Streams and rivers	0.0017	0.0001	0.0004	2–6 months
Biosphere	0.0006	0.00004	0.002	
Total	1,409			

(http://www.newworldencyclopedia.org/entry/Water_cycle)

Table 20.1-3 Applications of some cosmogenic radionuclides in hydrology

Nuclide	Half-life	Surface water	Ground water
³ H	12.3 years	Water age	Old-young
⁷ Be/ ¹⁰ Be	53 d/1.38 Myr	Particle settling	
¹⁴ C	5,730 years	Deep water formation	
³² Si	140 years	Dating	
³⁶ Cl	301 kyr	Bomb–prebomb comparison	Dating
⁸¹ Kr	229 kyr		Dating
¹²⁹ I	15.8 Myr	Ocean circulation	

residence time of the surface water in the mixed layer (Sect. 13.5.3) is much shorter than the figure given in Table 20.1-2. The deep water is subject to thermohaline circulation with a typical timescale of 2,000 years. Similarly, in the case of the Greenlandic ice sheets, the low-altitude regions near the coast break up into icebergs and melt every summer while in the central part the deepest ice is probably older than a million years.

Water is a very precious and valuable resource. Mankind is increasingly depleting the old water in the aquifers and placing major strains on the runoff in the streams and lakes. Increasingly we need to know where water has come from; how old it is; the location of the recharge zones of the aquifers; and other parameters that will allow us to manage this crucial resource with greater skill. The cosmogenic radionuclides provide vital tracing and dating tools in this challenge and in the following we illustrate their use in a number of applications. We will divide the applications into two groups relating to (1) surface waters (unsaturated zone, catchments, rivers, and lakes) and (2) ground water, and will discuss the potential and limitations of some cosmogenic radionuclides in these fields. We do not intend to provide a complete overview, but rather give the reader a flavour of what can be done. For more detailed discussions we refer to Michel et al. (2009).

Table (20.1-3) is a very condensed and incomplete summary of some applications.

20.2 Tritium

Tritium (T, ^3H) has some properties that make it an almost ideal tracer in hydrological applications. First of all it occurs as the “tritiated water” molecule ($^3\text{H}_2\text{O}$, sometimes written as HTO) and hence its geochemical behaviour is identical to that of H_2O . Second, in addition to its natural origin from cosmic ray-induced spallation reactions in the atmosphere (Sect. 10.3), it was produced in large amounts during the nuclear bomb test period from 1950 to 1963, generating a very strong and highly variable input signal (Fig. 12.3.1-2). Third, it decays into ^3He which not only assists the detection of Tritium (T) in very low concentrations (Sect. 15.1), but also provides information about when a water body was last in contact with the atmosphere, as we outline below.

The short half-life of 12.3 years means that the bomb pulse has largely decayed at the time of writing (2010), and therefore the role of T in hydrology is now reduced compared to the recent past. To some extent T can be replaced by ^{36}Cl (half-life: 301,000 years) as far as applications that use the bomb peak are concerned.

The main application of T in *surface waters studies* is in identifying and quantifying the physical processes that are in operation in the region under investigation. To this end, the T output signal from a watershed, lake, or reservoir is compared with the known T input signal. This provides information about mixing ratios between old (prebomb) and young (after bomb) water, reservoir sizes, and residence times. Since HTO participates in all processes such as evaporation, percolation, condensation, and infiltration in the same way as H_2O , it represents an optimal tool to develop and calibrate hydrological models. An example of such an application is given by the tritium balance of a lake:

$$\frac{dC_l}{dt} = -\lambda C_l + PC_p + RC_r + G_{in}C_g - G_{out}C_l - OC_l - E \quad (20.2-1)$$

where λ is the Tritium decay constant, l, p, and g stand for lake, precipitation, and groundwater, respectively, and C denotes the corresponding concentrations. P , R , and G_{in} are the positive mass flow rates into the lake from precipitation, catchment runoff, and from groundwater, respectively. G_{out} , O , and E are the negative mass flow rates through loss into the groundwater, surface outflow, and evaporation. Measurement of the various HTO concentrations provides the ability to investigate the overall balance between the several input and output mass flow rates.

Lakes are usually stratified with the denser (colder) water at the bottom and the less dense (warmer) water at the surface. During winter, surface cooling can cause a density inversion that leads to an overturn of the water with important consequences for the biogeochemistry and the ecology of the lake. Measuring T and ^3He in a water parcel provides an elegant way to determine when this parcel was last in contact with the atmosphere. Thus a water parcel at the surface is in exchange with the atmosphere and its He content equilibrates to the atmospheric value.

When surface water is transported deeper down in the lake the ^3He concentration increases as a result of tritium decay.

$$\frac{d\text{He}}{dt} = \lambda T(t) = \lambda T_0 e^{-\lambda t} \quad (20.2-2)$$

where λ is the decay constant of Tritium. $T(t)$ can be replaced by $T_0 e^{-\lambda t}$ with the Tritium concentrations T_0 at initial submersion and $T(t)$ at a later time t . Integrating over time, the ^3He concentration formed by the decay of T at time τ is given by

$$^3\text{He} = T_0(-e^{-\lambda\tau} + 1) \quad (20.2-3)$$

Replacing T_0 by $T(\tau)$ from (20.2-2) leads to

$^3\text{He} = T e^{\lambda\tau}(-e^{-\lambda\tau} + 1)$, hence

$$\tau = \frac{1}{\lambda} \ln \left(1 + \frac{^3\text{He}}{T} \right) \quad (20.2-4)$$

where τ is the time since the last gas exchange of the water parcel with the atmosphere. Thus the measurement of the $^3\text{He}/T$ ratio in a water sample directly yields an estimate of τ .

As noted above, the surface water contains a very small quantity of ^3He which is to be added to that in (20.2-3), and for which correction can be made in (20.2-4) (Schlosser et al. 1988).

Because of its short half-life T can only be applied to young *ground water*. The applications are similar to the ones described for surface waters. The radioactive decay properties can be used to determine flow rates and residence times while measurement of the $T/^3\text{He}$ ratio provides assistance in estimating recharge rates (Edmunds et al. 2006).

20.3 Carbon-14

Although ^{14}C is a powerful tool in the study of the movement of ocean water (McNichol and Aluwihare 2007) its applicability to the study of *surface and groundwater* is rather limited. This is mainly because of the distortions introduced by the exchange processes that occur between the water and the carbonates in soil and rocks, and to a lesser extent with CO_2 in the atmosphere. The carbonates are very old and hence free of ^{14}C , and this makes the water seem older than it is. (This is called the hard water effect.)

The same is true for *lake water*. This causes problems when dating sediments based on organic matter grown in the lake. Carbonate dissolution reduces the $^{14}\text{C}/^{12}\text{C}$ ratio and therefore increases the radiocarbon ages. The interpretation of radiocarbon in hydrogeology is further complicated by isotopic exchange and fractionation processes (Geyh et al. 1998).

20.4 Krypton-81

^{81}Kr is the most promising cosmogenic radionuclide for dating *old groundwater*. Its half-life of 229 kyr is similar to that of ^{36}Cl and it provides a dating range from about 50 kyr to 1–2 million years. It has several properties which make it to an almost ideal dating tool:

- The atmospheric $^{81}\text{Kr}/\text{Kr}$ ratio is the same all over the globe and well known.
- The atmospheric $^{81}\text{Kr}/\text{Kr}$ ratio can be considered to be constant over the past million years because all the effects of solar and geomagnetic modulation will have been averaged out over time.
- The input $^{81}\text{Kr}/\text{Kr}$ ratio in recharge areas of aquifers can be considered as spatially and temporarily constant.
- In-situ production within aquifers is almost negligible.

There are two difficulties, however. (1) Kr is a rare element in nature (Table 10.1-1), and therefore the production rate of ^{81}Kr is low (see below). Consequently its concentration in water is very low. (2) As the noble gases do not form negative ions, ^{81}Kr cannot be measured with the standard AMS technique based on Tandem accelerators and requires the use of cyclotron or laser techniques (Sect. 15.5).

^{81}Kr is produced almost exclusively by $^{80}\text{Kr}(n,\gamma)^{81}\text{Kr}$ and by spallation of the heavier isotopes of Kr (82 to 86). Anthropogenic production due to nuclear bomb tests can be neglected. The only known additional source is spontaneous fission of ^{238}U underground. However, ^{81}Kr is shielded by the stable nuclide ^{81}Br which means that the β -decay chain stops before reaching ^{81}Kr (Lehmann et al. 1993).

The estimated atmospheric $^{81}\text{Kr}/\text{Kr}$ ratio is about 5×10^{-13} . This corresponds to a total global inventory of 6.4×10^{25} ^{81}Kr atoms. Gas exchange with water at 15°C leads to a concentration of about 1,000 ^{81}Kr atoms in 1 l of water. ^{81}Kr is much less soluble in water than is CO_2 and consequently only about 2% of the global inventory of ^{81}Kr is stored in water while 98% resides in the atmosphere. As a consequence changes in ocean circulation do not affect the atmospheric concentration of ^{81}Kr , a clear difference from the case of ^{14}C .

A good example of the ^{81}Kr dating technique is provided by a study carried out in the Great Artesian Basin (GAB) of Australia (Lehmann et al. 2003). The GAB is the largest and deepest artesian basin on the globe and covers 1.7 million square kilometres, this being about a quarter of Australia (Fig. 20.4-1). It extends to a depth of about 3 km and is estimated to contain some 65,000 km³ of water which equals more than half the global annual precipitation over land and represents a major source of water for the desert regions of inland Australia.

Ground water samples were taken from the four wells indicated on the map: Oodnadatta (OD, 1), Raspberry Creek (RC, 2), Watson Creek (WC, 3), and Duck Hole (DH, 4). The ages were calculated assuming radioactive decay with the $^{81}\text{Kr}/\text{Kr}$ ratio of the recharge area taken as the initial value (Fig. 20.4-2). The analytical uncertainty is about 10–20%. Additional uncertainties were estimated leading to the dashed lines in Fig. 20.4-2. Overall, this demonstrates rather well the type of results that can be obtained using ^{81}Kr for dating old ground water.

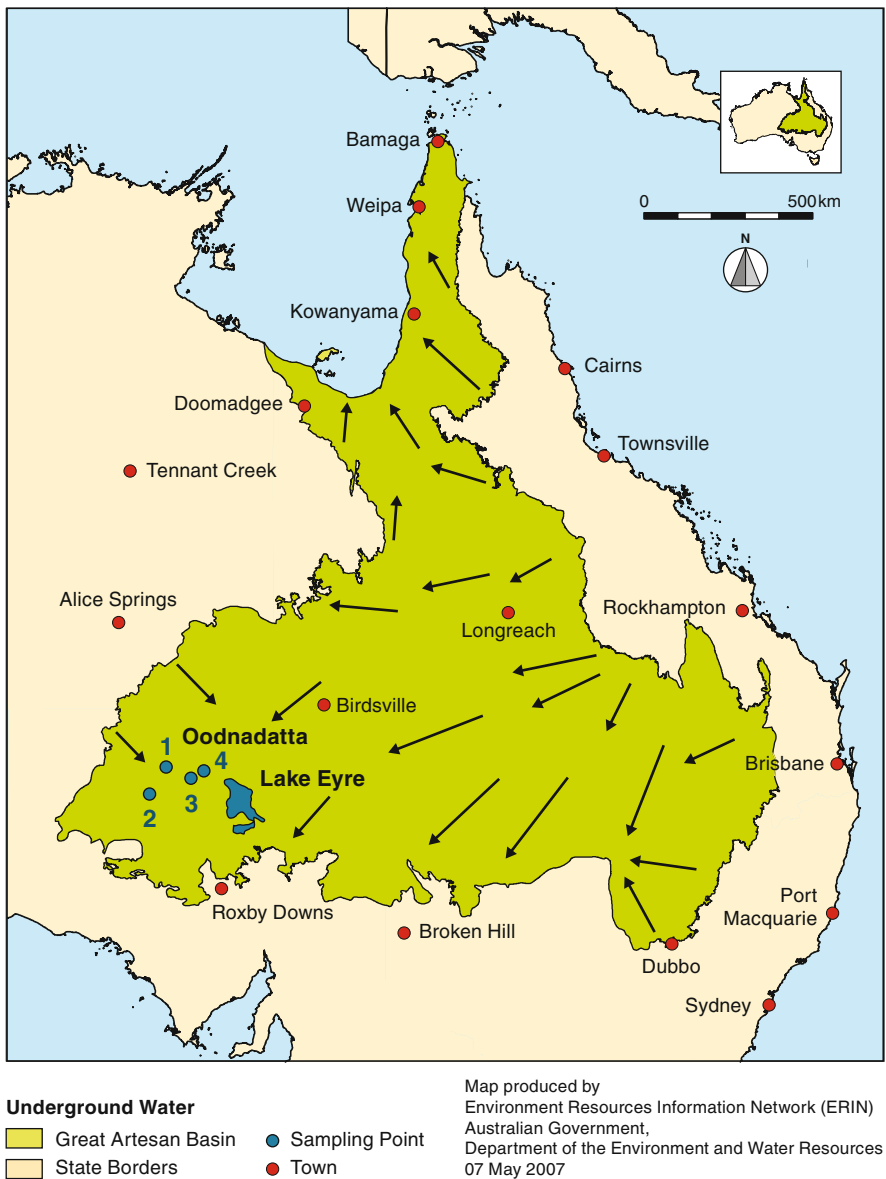
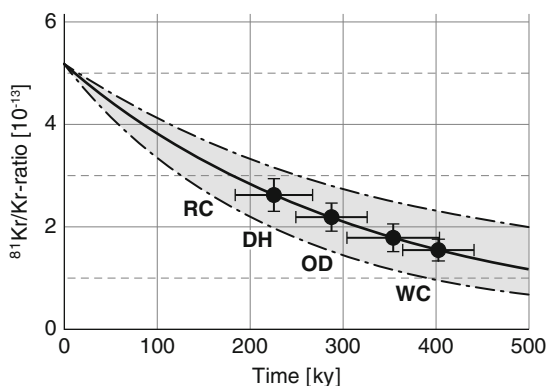


Fig. 20.4-1 Great Artesian Basin (GAB) with the four sampling sites Oodnadatta (1), Raspberry Creek (2), Watson Creek (3), and Duck Hole (4). The general flow directions of the ground water are indicated by *arrows*. After Collon et al. (2000)

Fig. 20.4-2 $^{81}\text{Kr}/\text{Kr}$ ratio as a function of time for the four sampling sites assuming radioactive decay only. After Lehmann et al. (2003)



20.5 Chlorine-36

As already mentioned ^{36}Cl has similarities to ^3H in the sense that for both radionuclides the natural cosmic ray-induced production rate was strongly overprinted by the nuclear bomb tests between 1950 and 1963. Because of its different production mechanism (neutron activation of ^{35}Cl in seawater; see Sect. 19.3) the ^{36}Cl peak appeared earlier (1958/59) than did the T peak (1963/64), and was produced by a relatively small number of large nuclear tests on atolls and ships. The dates and sites of these explosions are well known. Unfortunately, this is not the case for the exact amount of ^{36}Cl which was injected into the stratosphere which has therefore to be estimated based on the magnitude of the explosions. Modelling using the ECHAM5-HAM Global circulation model reveals a generally good agreement with the ^{36}Cl profiles measured in polar ice sheets and mountain glaciers, as outlined in Sect. 19.3. Figure 19.3-3 shows a comparison of the ^{36}Cl fluxes derived from measurements on eight ice cores distributed over the globe and the model calculations based on the available data on the nuclear bomb tests (Heikkilä et al. 2009).

The temporal evolution of the modelled latitudinal deposition (integrated over all longitudes) is shown in Fig. 20.5-1. It shows that the largest deposition occurred in the Northern Hemisphere between 1954 and 1960. The tests around 1970 were carried out by the French in the Southern Hemisphere with very little effect on the Northern Hemisphere. Detailed maps of the ^{36}Cl deposition fluxes are used to provide the input data for hydrological applications.

There was a great hope that ^{36}Cl would be another perfect dating tool for very old groundwater. Its half-life of 301,000 years and its high solubility are ideal for dating groundwater in the range from 0.1 to 2 million years. However, experience has shown that there are several difficulties, the main ones being:

There is substantial underground production of ^{36}Cl by neutron activation of dissolved ^{35}Cl ($^{35}\text{Cl}(n,\gamma)^{36}\text{Cl}$) in the aquifer. The neutron flux originates in spontaneous decay of the Uranium and Thorium in the rocks in the aquifer, and is further modified by the presence of neutron-absorbing elements. Additional production takes place within the aquifer, for example, by muon reactions with carbonate

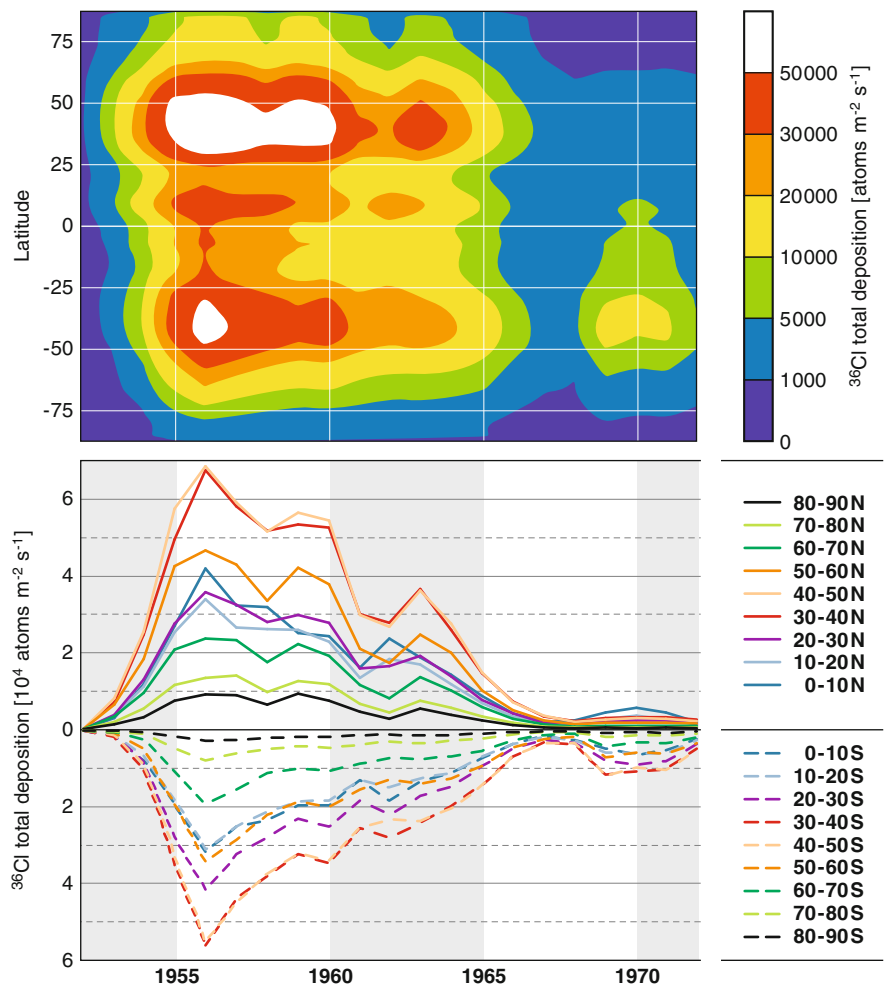


Fig. 20.5-1 Longitudinally integrated, modelled ^{36}Cl deposition fluxes as a function of latitude. After (Heikkila et al. 2009)

($^{40}\text{Ca}(\mu, \alpha)^{36}\text{Cl}$) (Sect. 11.2.1). Dissolution of “old” chlorine from the rocks in the aquifer leads to a reduction of the $^{36}\text{Cl}/\text{Cl}$ ratio.

The combination of radioactive decay, underground production, and chloride dissolution results in the following equation:

$$RC = R_0C_0e^{-\lambda t} + R_{\text{eq}}C_0(1 - e^{-\lambda t}) + R_{\text{eq}}(C - C_0) \quad (20.5-1)$$

where R and C are the measured $^{36}\text{Cl}/\text{Cl}$ ratio and chloride concentration in the groundwater, R_0 and C_0 are the corresponding initial values in the recharge area of the aquifer, and R_{eq} is the $^{36}\text{Cl}/\text{Cl}$ ratio at secular equilibrium in the underground.

In the case of the GAB (Fig. 20.4-2) two of the four samples provide ^{36}Cl ages in good agreement with ^{81}Kr (Raspberry Creek and Oodnadatta) while for the other two underground processes resulted in substantial discrepancies.

In a study of groundwater from the Sahara, where ^{81}Kr and ^{36}Cl dating yielded consistent results, the conditions were more favourable (Sturchio et al. 2004). Figure 20.5-2 shows the results obtained from six samples from the Nubian aquifer in Egypt.

Assuming a constant chloride concentration $C = C_0$ Eq. (20.5-1) is reduced to:

$$R = R_{\text{eq}} + (R_0 - R_{\text{eq}})e^{-\lambda t_{\text{Kr}}} \quad (20.5-2)$$

with t_{Kr} being the ages determined by ^{81}Kr . Figure 20.5-2 shows that the ^{81}Kr ages are reasonably well fitted by this equation. The intercept with the y-axis gives the initial $^{36}\text{Cl}/\text{Cl}$ ratio R_0 in the recharge area and its value of $131 \pm 11 \times 10^{-15}$ is in good agreement with the value estimated for the GAB. The secular equilibrium value R_{eq} turns out to be $8 \pm 3 \times 10^{-15}$. The Sherka sample is the only one where there is significant disagreement between the ^{81}Kr and ^{36}Cl ages; it had a $^{36}\text{Cl}/\text{Cl}$ ratio that is considerably less than that predicted by the ^{81}Kr (6.8×10^5 years) age. This implies an age of 1.5×10^6 years and may be due to the dissolution of “old” chloride from the rocks in the aquifer.

These examples show that dating old groundwater is subject to a number of errors, and that it can only be done reliably by applying different techniques and measuring several parameters which provide a better insight into the different processes taking place. However, in view of the great importance of groundwater these attempts are justified and further improvements in the ^{81}Kr -measuring technique are expected.

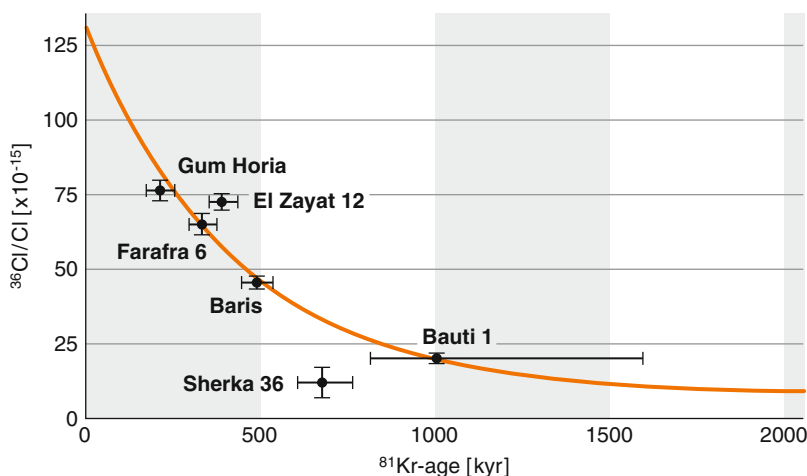


Fig. 20.5-2 Comparison of the $^{36}\text{Cl}/\text{Cl}$ ratio with the ^{81}Kr ages for the groundwater samples from the Nubian aquifer (Sturchio et al. 2004). The red curve shows the best exponential fit for ^{36}Cl based on Eq. (20.5-2)

20.6 Beryllium-7 to Beryllium-10 Ratio

As discussed before, these two cosmogenic radionuclides have very different half-lives which allow them to be used to investigate the rates at which processes occur in nature. This has important applications in hydrology, as illustrated by the following example.

An important difference between lakes and oceans is that, as a consequence of their smaller size, boundary effects play a central role in the hydrology of lakes. Thus adsorption, coagulation, and boundary uptake and release of trace elements and radionuclides affect the particle settling processes and the residence times in lakes.

To illustrate the use of the $^7\text{Be}/^{10}\text{B}$ ratio, Fig. 20.6-1 shows an experimental set-up in Lake Zurich where two sediment traps were deployed at water depths of 50 and 130 m. Lake Zurich is a U-shaped lake with steep lateral slopes and consists of two basins with a maximum depth of 137 m.

During the period 1983–1987 the sediment traps were deployed for time intervals ranging between 2 and 3 weeks. The collected particles were analysed for ^7Be , ^{10}Be , ^{210}Pb , and ^{137}Cs (Wieland et al. 1991). In the following we concentrate on the ^7Be and ^{10}Be results.

As the main inflow comes from another lake about 10 km upstream most of the ^7Be and ^{10}Be had already been removed and can be neglected. The dominant source is therefore the input by precipitation directly into the lake. Figure 20.6-2 shows the ^7Be and ^{10}Be concentrations for the particles collected in the two sediment traps in the year 1987.

The ^7Be concentrations in trap B are generally smaller than in trap A. This is not the case for ^{10}Be . The obvious implication is that some of the ^7Be has decayed

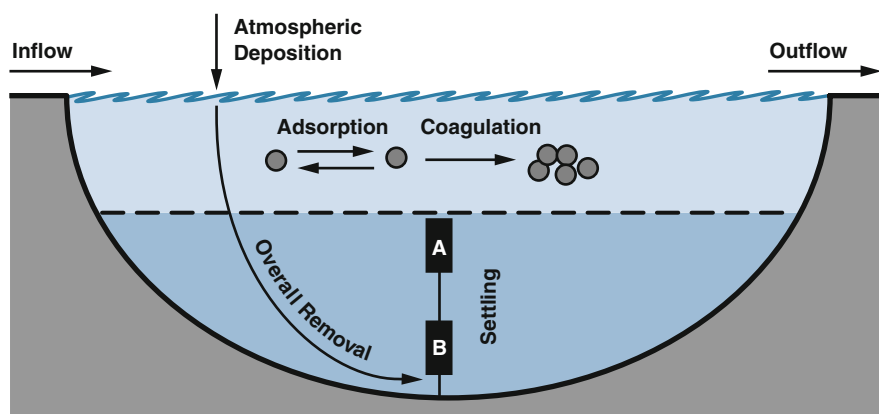


Fig. 20.6-1 The experimental set-up in Lake Zurich used to measure the $^7\text{Be}/^{10}\text{B}$ ratio as a function of depth, and thence the particulate settling velocity. A and B represent sediment traps at 50 and 130 m depth, respectively. After (Wieland et al. 1991)

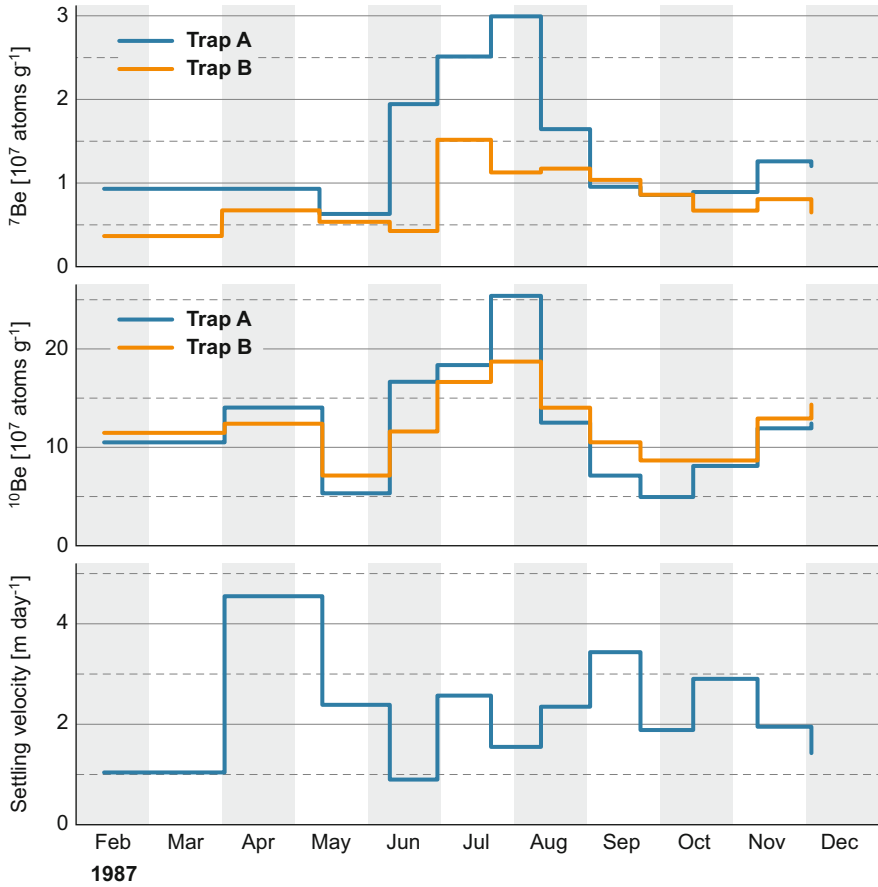


Fig. 20.6-2 ^7Be and ^{10}Be concentrations (*upper and middle panel*) of particles collected by trap A at 50 m depth and trap B at 130 m depth in Lake Zurich during the year 1987. The third panel gives the settling velocity computed from those data (Wieland et al. 1991)

during settling. To correct for boundary processes we normalize ^7Be to ^{10}Be and obtain

$$\frac{^7\text{Be}}{^{10}\text{Be}}(\text{B}) = \frac{^7\text{Be}}{^{10}\text{Be}}(\text{A}) \cdot e^{-\lambda \cdot \Delta t} \quad (20.6-1)$$

where λ is the ^7Be decay constant, and Δt is the settling time between traps A and B. The settling velocity can be derived from Δt in (20.6-1) and the distance between the sediment traps, as displayed in the third panel of Fig. 20.6-2. In this example, the average settling velocity is about 2 m per day, while the settling time Δt is ~ 40 days.

References

- Collon P et al (2000) Kr-81 in the Great Artesian Basin, Australia: a new method for dating very old groundwater. *Earth Planet Sci Lett* 182(1):103–113
- Edmunds WM, Ma JZ, Aeschbach-Hertig W, Kipfer R, Darbyshire DPF (2006) Groundwater recharge history and hydrogeochemical evolution in the Minqin Basin, North West China. *Appl Geochem* 21(12):2148–2170
- Geyh MA, Schotterer U, Grosjean M (1998) Temporal changes of the C-14 reservoir effect in lakes. *Radiocarbon* 40(2):921–931
- Heikkilä U, Beer J, Feichter J, Alfimov V, Synal HA, Schotterer U, Eichler A, Schwikowski M, Thompson L (2009) Cl-36 bomb peak: comparison of modeled and measured data. *Atmos Chem Phys* 9(12):4145–4156
- Lehmann BE, Davis SN, Fabryka-Martin JT (1993) Atmospheric and subsurface sources of stable and radioactive nuclides used for groundwater dating. *Water Resour Res* 29(7):2027–2040
- Lehmann BE et al (2003) A comparison of groundwater dating with Kr-81, Cl-36 and He-4 in four wells of the Great Artesian Basin, Australia. *Earth Planet Sci Lett* 211(3–4):237–250
- McNichol AP, Aluwihare LI (2007) The power of radiocarbon in biogeochemical studies of the marine carbon cycle: Insights from studies of dissolved and particulate organic carbon (DOC and POC). *Chem Rev* 107(2):443–466
- Michel RL, Kraemer T, Cecil LD (2009) Chapter 5 Radionuclides as tracers and timers in surface and groundwater. In: Klaus F (ed) *Radioactivity in the environment*. Elsevier, Amsterdam, pp 139–230
- Schlosser P, Stute M, Dörr H, Sonntag C, Münnich KO (1988) Tritium/³He dating of shallow groundwater. *Earth Planet Sci Lett* 89:353–362
- Sturchio NC et al (2004) One million year old groundwater in the Sahara revealed by krypton-81 and chlorine-36. *Geophys Res Lett* 31:L05503
- Wieland E, Santschi PH, Beer J (1991) A multi-tracer study of radionuclides in Lake Zürich, Switzerland: 2. Residence times, removal processes, and sediment focusing comparison of atmospheric and sedimentary fluxes of ⁷Be, ¹⁰Be, ²¹⁰Pb, ²¹⁰Po, and ¹³⁷Cs. *J Geophys Res* 96:17067–17080

Chapter 21

Geosphere

21.1 Introduction

The Earth was formed 4.5 billion years ago together with the Sun and the other planets of the solar system and as a consequence their original compositions were the same. High temperatures caused the smaller inner planets to lose a large percentage of their gases and they are therefore often called the stony planets. Gravitational energy released by the compaction of planetary mass and the decay of primordial radioactive nuclides (^{40}K , ^{235}U , ^{238}U , and ^{232}Th) caused melting and separation of planetary matter according to density.

The interior of Earth is divided into five layers (Fig. 21.1-1). The crustal thickness varies from 30–50 km on the continents to typically 6 km under the oceans and consists mainly of silicates. The crust together with the top part of the mantle is called the lithosphere. It “floats” on the viscous mantle beneath and gives rise to the motion of the tectonic plates and associated phenomena such as volcanic eruptions and earthquakes. Analysis of the propagation of the “seismic” waves from earthquakes has shown that the outer core is in a liquid state, and contains large amounts of iron. Convection and rotation in this layer forms a dynamo which generates the geomagnetic dipole field. The inner core is solid.

For a long time it was believed that the geosphere was a static system in strong contrast to the very dynamic atmosphere. However, this impression was wrong. While the atmosphere changes within minutes to hours, the processes in the geosphere occur on “geological” time scales much longer than a person’s lifetime and were therefore not obvious to science until the nineteenth century. As already mentioned, the energy sources within the geosphere are of gravitational (20%) and nuclear (80%) origin, and amount to 4×10^{13} W, corresponding to about 40,000 nuclear power plants.

For engineering, hydrological, agricultural, and many other reasons, modern society finds it important to understand the processes that have fashioned and

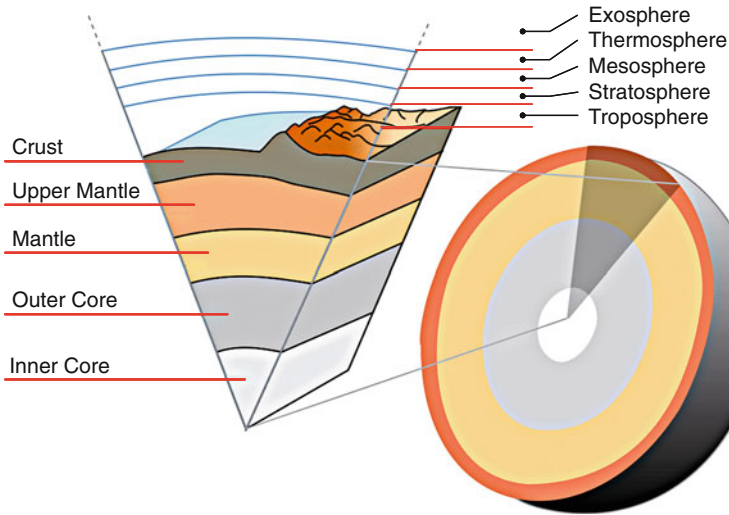


Fig. 21.1-1 Internal structure of Earth (Wikimedia Commons)

are continuing to alter the landforms. A number of these natural geomorphological processes are listed below; the cosmogenic radionuclides have important applications in the study of many of them. In recent years, anthropogenic activities have played an increasingly important role in this field (e.g. in the initiation of erosion of soil) and here too the cosmogenic nuclides have important applications.

Tectonic Processes. Tectonic processes are responsible for the motion of the continental plates, the distribution of the continents, the formation of mountain chains and volcanoes, the subduction of ocean sediments under continental plates, and the occurrence of earth quakes.

Glacial. The effects of glaciers and the large ice sheets during the glacial epochs have been large, but mainly restricted to cold areas. The movement of the ice has widened valleys, formed moraines and lakes, and generated a great deal of fine-grained material. Debris and large boulders have been transported during glacial periods over long distances creating “erratic” boulders.

Hill Slope. Gravity causes all mass to move downwards, and this has played a part in shaping terrestrial and submarine slopes.

Fluvial. Water flowing in rivers and streams transports matter picked up in the catchment area or mobilized from the sediments. This matter can be in particulate or dissolved form.

Aeolian. Although less effective than water, wind also contributes to the reshaping of the landscape. Especially in dry unconsolidated areas such as deserts, large amounts of dust are picked up by winds and transported over long distances to form loess plateaus (Sect. 14.4).

Cosmogenic radionuclides can be applied to most of the above-mentioned processes to study either the process itself, or to gain information about its dynamics through the use of radioactive decay to measure the age of deposits, etc.

21.2 Geomagnetic Field Intensity

More than 1,000 years ago the Chinese discovered that a needle-shaped piece of lodestone, a naturally magnetized piece of the mineral magnetite, always aligns itself in the north–south direction when floating in a bowl of water. This led to the invention of the compass, an instrument of fundamental importance for the development of modern civilisation. They were soon in worldwide navigational use. Navigators reported somewhat perplexing observations, which William Gilbert explained in 1600 by describing the Earth as a giant magnet. Two hundred years later Gauss developed a mathematical model showing that the source of the majority of the measured magnetic field is inside the Earth. Later it became clear that the field is produced by a self-sustained dynamo process driven by the rotation and convection of the electrically conducting molten material in the outer core of the Earth. This is the same type of process which is responsible for the magnetic field of the Sun and the other stars.

As discussed in Sects. 5.7 and 5.8 solar activity and the strength of the geomagnetic dipole field are the two main factors that control the cosmic ray flux reaching the Earth. It is possible (but not yet proven) that there have been long-term changes in the flux over time scales of 10,000 years and greater as a result of changes in the properties of the interstellar medium near Earth; variations across the spiral arms of the galaxy; and the occurrence of nearby supernova (Chap. 18). Apart from these the vast majority of all production changes of the cosmogenic radionuclides are the result of either solar or geomagnetic modulation of the galactic cosmic ray flux (Sect. 10.3).

Detailed measurements of the present-day geomagnetic field show that it can be regarded as a dipole field, off-set from the centre of the Earth by ~ 400 km, upon which are superimposed higher-order components which change as functions of space and time (Sect. 5.8). All of the higher-order components of the geomagnetic field vary as a function of longitude, Φ , in the form of $\sin(n\Phi)$, where n is an integer (Eq. 5.8.1-1). To a close approximation then, the cut-off rigidity, and consequently the cosmogenic production rate at a given geographic latitude vary sinusoidally as we proceed around the world (i.e. $0 < \Phi < 360$). That is, they vary (by a relatively small amount) above and below mean values determined by the Stoermer cut-off for that latitude (Eq. 5.8.2-2). Atmospheric mixing (Chap. 13) will average out these variations in both the stratosphere and the troposphere, with the result that the dipole component (and the Stoermer equations 5.8.2-1 and 5.8.2-2) is the primary determinant of the worldwide variation in the cosmogenic production rates. Polar wander (of the dipole – Fig. 5.8.1-1) introduces minor changes to the above argument that result in changes of $< 7.5\%$ in the ^{10}Be , say, entering the polar ice archive (McCracken 2004).

The dipole field reduces the cosmic ray flux preferentially at low latitudes where the particles impinge at right angles to the horizontal field lines and therefore experience the greatest deflecting forces. At geomagnetic latitudes above 65 degrees the Stoermer cut-off approaches zero and cosmic rays of all energies can reach the

Earth. The physical reason for this is that the magnetic field lines of the dipole extend far out into space, and they guide the cosmic ray particles of all energies to the top of the atmosphere. As a consequence, changes in the geomagnetic dipole field have the greatest effect on the production rate of cosmogenic radionuclides at low geomagnetic latitudes, whereas they have no effect at high latitudes. This is sometimes called “geomagnetic modulation”.

As discussed in Sect. 5.7, and illustrated in Figs. 5.7.3-1 and 10.2.3-1, solar modulation is greatest at low energies, and decreases steadily towards higher energies. The low-energy cosmic rays are prevented from reaching the atmosphere by the geomagnetic dipole field at low, but not at high, latitudes, and consequently solar modulation is particularly strong at high latitudes. Figure 21.2-1 compares the latitude dependence of the solar and the geomagnetic modulation on the production rate of ^{10}Be . A change of the solar modulation function Φ from 900 to 300 MeV with the geomagnetic dipole moment $M = 1$ leads to an almost identical quantitative change in the latitudinal production rate as a change of the dipole field from $M = 1.5$ to $M = 0.5$ with $\Phi = 550$ MeV. The comparison reveals the clear dominance of geomagnetic modulation below 30° and the dominance of solar modulation above 60° . Between 30° and 60° the gradients of both modulation effects are steep. This means that any change of the position of the magnetic poles (polar wandering, see Fig. 5.8.1-1) causes modulation effects in this latitude band even if the cosmic ray intensity remains constant. For example, polar wander between 78° and 90° (as evident over the past 1,000 year) is computed to result in $\sim 5\%$ to 7.5% change in ^{10}Be entering the polar archive.

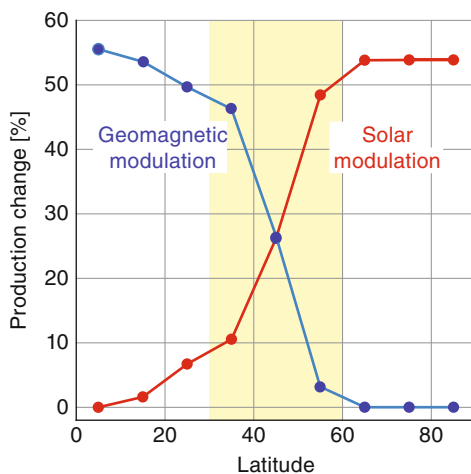


Fig. 21.2-1 Illustrating the differing nature of solar and geomagnetic modulation of the galactic cosmic radiation. Relative changes in the ^{10}Be production rate are shown versus latitude. The *blue* curve depicts the relative production rate for a change of the geomagnetic dipole moment from 1.5 to 0.5 of the present value. The *red* curve shows the corresponding change when the solar modulation function Φ drops from 900 to 300 MeV. Both forms of modulation are significant in the coloured region $30 < \theta < 60$

At first sight, the fundamentally different latitude dependences shown in Fig. 21.2-1 appear to provide the means to decide whether observed changes in the concentration (or flux) of the cosmogenic radionuclides have been induced by solar or geomagnetic modulation. However, as shown in Sect. 13.4 and Fig. 13.4-2, the latitudinal production pattern is completely changed by the transport of the cosmogenic radionuclides from the atmosphere into the respective archives. Thus while Fig. 21.2-1 shows that there is no geomagnetic modulation of the production in the polar region, atmospheric mixing means that some of the equatorial production (with its strong imprint of geomagnetic modulation) will be sequestered in the polar archive. In the same manner, any record of a cosmogenic radionuclide is a priori a mixture of solar and geomagnetic production changes which have been further modified by the transport processes.

The situation is quite different in the case of in-situ production. Depending on the geomagnetic latitude, the in-situ production mainly reflects magnetic or solar modulation. However, the in-situ produced radionuclides reflect the time integral of the production rate, and are therefore not suited to the study of temporal variability. The only potential exception is the determination of long-term mean values, such as in the case of meteorites.

From the above, it is clear that atmospheric mixing makes it almost impossible to separate the solar and geomagnetic signals in the tree ring or ice archives. The only physical argument that allows separation of the two effects is the inertia of the Earth dynamo, which is unlikely to change significantly on time scales shorter than a few centuries, in contrast to solar variability which is characterized by the very strong 11-year Schwabe cycle and major 50–100 year events such as the Grand Minima. This argument is supported by the fact that a reversal of the geomagnetic dipole (as recorded in the paleomagnetic record, Sect. 5.8.1) typically lasts a few thousand years. Likewise, the secular changes in the magnetic dipole moment (e.g. Fig. 5.8.1-2) occur over intervals of ~1,000 years, while significant polar wander (Fig. 5.8.1-1) occurs over intervals of 500 years or so. Therefore, it has been a common assumption that production changes on decadal to centennial time scales are of solar origin and changes on multi-millennial time scales are the consequence of geomagnetic dipole effects. These assumptions seemed reasonable in the past when we had relatively short cosmogenic records (<1,000 year), but the acquisition of 10,000-year high-quality ^{14}C and ^{10}Be records (e.g. Fig. 7.3-3) indicates that the Sun exhibits variability on multi-millennial time scales as well. For this reason, it is now common practice to use the paleomagnetic record to compensate for magnetic changes prior to investigating solar effects. As outlined in Sect. 5.8.1, the traditional way to study the paleomagnetic field has been to measure the magnetic properties of sediments, lava, or pottery. For a certain time after settling to the bottom of a lake, magnetic particles are still mobile and they align themselves with the local magnetic field lines. After burial under new layers of sediment, they become immobile at a certain depth, the so-called lock-in depth, and they thereafter store the direction and the intensity of the local field prior to “lock-in”. In lava or pottery the geomagnetic field is stored when the temperature drops below the Curie point. Applying sophisticated methods it is possible to determine the direction (declination and inclination) and the

intensity of the archive field after many thousand years (McElhinney and McFadden 2000). As outlined in Sect. 5.8.1, however, geological, chemical and biological effects may overprint or modify the stored signal. Another problem is that the non-dipole components of the geomagnetic field are faithfully recorded along with the dipole. Nevertheless, with a growing number of paleomagnetic records distributed over a large part of the globe, the quality and the temporal resolution are continuously improving.

The cosmogenic radionuclides provide a second method to investigate the paleomagnetic field. Figure 21.2-2 shows the mean global production rate of ^{14}C (blue curves) and ^{10}Be (red curves) in relative units for different levels of solar activity ($\Phi = 100, 550, 1,000 \text{ MeV}$). For easier comparison the production rates were normalized to 1 for the present day dipole field intensity ($M = 1$) and a solar modulation function $\Phi = 550 \text{ MeV}$.

It is clear from Fig. 21.2-2 that the effect of a field change on the production rate is largest for small fields. Further, ^{14}C is more sensitive to geomagnetic modulation than ^{10}Be and most other cosmogenic radionuclides (not shown). The reason for this is that ^{14}C is produced by thermal neutrons which exhibit stronger modulation effects than the fast neutrons which are mainly responsible for high-energy spallation reactions that generate ^{10}Be .

From the above discussion it is obvious that the use of cosmogenic radionuclides to reconstruct the history of the geomagnetic field has two main limitations: (1) as a result of radioactive decay, the time span which can be covered ranging between 5 and 10 half-lives depending on the radionuclide and the analytical procedures used, and (2) the need to distinguish between production changes induced by solar and geomagnetic modulation.

Fortunately the strengths and the weaknesses of the two completely independent remanence and cosmogenic methods are largely complementary. While cosmogenic radionuclides are most sensitive for weak fields (Fig. 21.2-2) the remanence

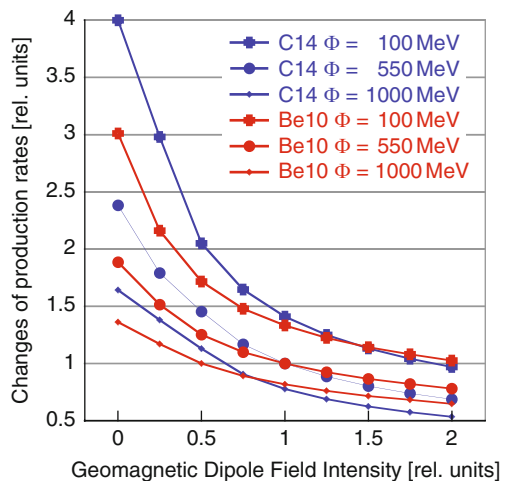


Fig. 21.2-2 Relative changes of the production rates of ^{14}C and ^{10}Be as a function of the geomagnetic dipole moment for different levels of solar activity and global atmospheric mixing. The results are based on the calculations of Masarik and Beer (2009)

method works best for strong fields. To illustrate this, analyses of the Laschamp event ($\sim 40,000$ years BP) and the Brunhes/Matuyama reversal that occurred $\sim 780,000$ years ago are outlined in the following.

Figure 21.2-3 shows the comparison of two reconstructions of the paleomagnetic field intensity expressed as the virtual axial dipole moment. The first is derived from ^{10}Be ice core data from Greenland, the second (called “NAPIS75”) from remanent magnetization data compiled from six sediment cores drilled in the North Atlantic (Laj et al. 2000). Panel (a) shows the ^{10}Be flux calculated from the ^{10}Be concentrations measured in the GRIP ice core from central Greenland. A gap between 10,000 and 17,000 BP was supplemented by ^{10}Be data from the nearby GISP core. For more details we refer to the study of Muscheler et al. (2005). To remove solar variability the data were low-pass filtered with a cut-off period of 3,000 years (grey band). These data were then converted to magnetic dipole

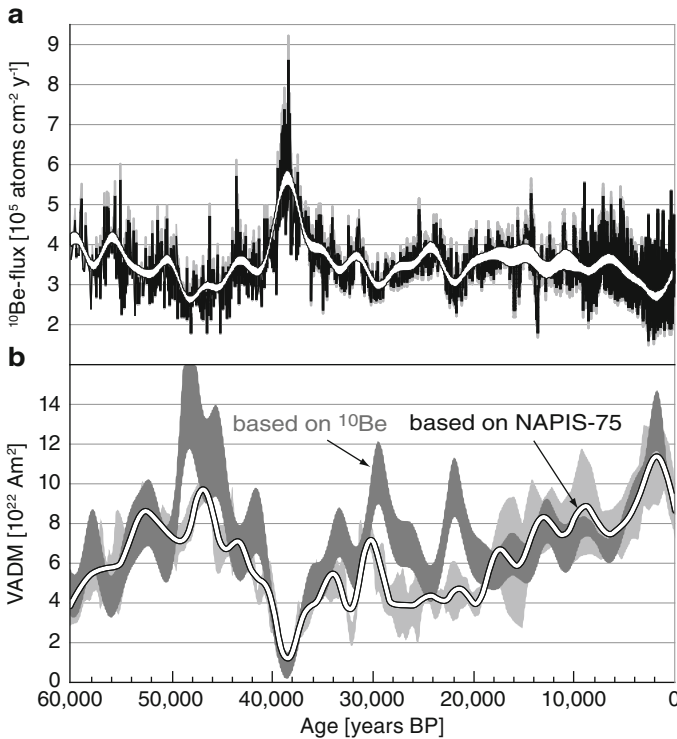


Fig. 21.2-3 Comparison of the intensity of the calculated virtual axial dipole moment (VADM) derived from ^{10}Be flux records in Greenland with the VADM record measured in sea sediment cores. Panel (a) shows the ^{10}Be flux from the GRIP ice core supplemented by GISP data in the interval 10000–20000 years BP and its long-term trend after applying a 3,000-year low-pass filter (grey band). Panel (b) is a comparison of the VADM record derived from the ^{10}Be flux in panel (a) with the NAPIS-75 paleointensity record stacked together from six north Atlantic sediment cores (Laj et al. 2000). Note the Laschamp excursion at about ~ 40 kyr BP [figure after Muscheler et al. (2005)]

moment as shown in Fig. 21.2-2, based on the assumption that the changes in ^{10}Be flux were because of geomagnetic changes alone. The result is shown in panel (b) as a dark grey band indicating the uncertainty of the reconstruction. The white line with a light grey band depicts the NAPIS-75 record derived from measurements of the remanent magnetizations.

Bearing in mind the various uncertainties and assumptions, there is good agreement between the two estimates of the geomagnetic variation shown in Fig. 21.2-3. These reconstructions show that, except for the time around 50,000 years BP, the geomagnetic dipole moment was generally lower than today ($8 \times 10^{22} \text{ A m}^2$), and that therefore the production rate of cosmogenic radionuclides was generally higher. Closer examination shows that there is reasonable agreement between the ~5,000-year-scale variations in both estimates. The Laschamp excursion at about 40,000 years BP is a striking feature of both records, each showing that the geomagnetic dipole moment attained values very close to zero. It seems that the geomagnetic field almost reversed. The ^{10}Be flux increased by a factor of ~2 as expected from the production calculations (Fig. 21.2-2 and Sect. 10.3). The same factor of 2 pertained after atmospheric redistribution (Fig. 13.4-4). The duration of the Laschamp excursion was about 2,000 years.

As discussed in Chap. 23 production peaks such as the Laschamp excursion can be used as time markers, either to synchronize records or, if the marker is dated, for absolute dating. It should be noted that a more recent determination of the Laschamp excursion using $^{40}\text{Ar}/^{39}\text{Ar}$, K-Ar, and $^{238}\text{U}/^{230}\text{Th}$ dating gave an age of $40,650 \pm 950$ years (Singer et al. 2009).

The last time the polarity of the earth's dipole moment reversed was 778 ± 2 kyr ago (Singer and Pringle 1996). This so-called Brunhes/Matuyama (B/M) reversal is well established from remanent magnetization measurements in sediments and lava flows. For a long time, however, it was too far in the past to be examined in ice cores. Recently, however, the new Dome C ice core drilled within the framework of EPICA (European Project for Ice Coring in Antarctica) provided a record which covers the necessary time span to investigate the B/M reversal.

Figure 21.2-4 shows the median ^{10}Be flux (Raisbeck et al. 2006) together with the corresponding inversely plotted geomagnetic field intensity from a compilation of more than 30 sedimentary records (Guyodo and Valet 1999). Only the time window covering the B/M reversal is shown. The reason for using the median flux is related to migration processes in the ice discussed in the next section. Going back in time the geomagnetic dipole moment was about the same as today ($8 \times 10^{22} \text{ A m}^2$) between 720 and 760 kyr BP. Then its magnitude dropped to almost zero at 780 kyr BP, the time at which the magnetization records indicate that the polarity reversal occurred. The corresponding calculated ^{10}Be production is shown as the dotted line in the upper panel. The data suggest that the duration of the B/M reversal was about 12 kyr, a considerably longer interval than in the case of the Laschamp excursion.

These two examples show that there is good overall agreement between the deductions based on the ^{10}Be flux and magnetic remanence measurements, indicating the validity of the assumptions involved, and that the ^{10}Be changes on these longer time scales were mainly because of geomagnetic modulation.

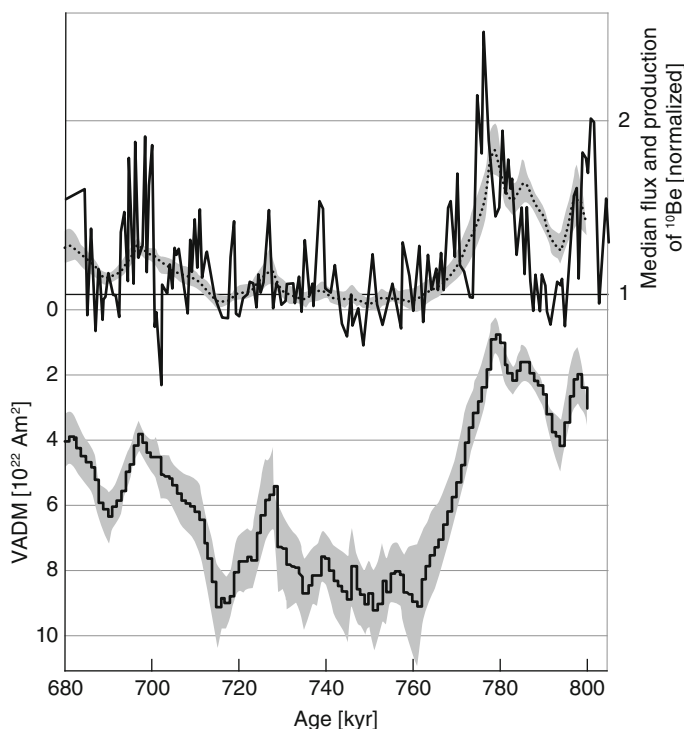


Fig. 21.2-4 Comparison of the median ^{10}Be flux (*upper panel solid line*) measured in the EPICA Dome C ice core with the expected ^{10}Be production rate (*upper panel dotted line*) derived from the inversely plotted geomagnetic field intensity (*lower panel solid histogram*) [figure after Raisbeck et al. (2006)]

21.3 Transport of Cosmogenic Radionuclides in Geological Systems

21.3.1 Introduction

All the subsystems of the terrestrial and space environment are dynamic, and it is natural to expect that there will be transport effects in the geosphere also, although generally on much longer time scales. Transport effects can be of major practical importance; for example, they are of vital concern if we are to store nuclear waste safely for millions of years in igneous or sedimentary formations. They can be applied to the study of the subduction of sea sediments under continental plates. On the other hand, transport effects can be a nuisance when they occur in ice cores which are expected to serve as an archive and to preserve the stored information in an undisturbed way over many millennia. The transport (migration) effects of the cosmogenic radionuclides in the geosphere depend strongly on their geochemical properties and their half-lives, and we discuss several examples in the following.

21.3.2 Migration in Ice

Ice has a reputation as an almost perfect archive. It collects everything removed from the atmosphere by wet and dry precipitation and keeps it deep-frozen for many millennia. It is well known that ice is viscous, and that it flows slowly, leading to a thinning of the annual ice layers at substantial depths. Nevertheless, it was a surprise when ^{10}Be measurements on ice samples older than 100 kyr gave different results depending on whether they were filtered or not before processing. One way to explain this unexpected result was to assume that over long periods of time ^{10}Be migrates within the ice and is concentrated on dust grains, which are then removed by the filter. In support of this, it has been observed that during recrystallization processes, impurities such as H_2SO_4 are swept out from the lattices into the boundaries. Extensive tests have shown that typically 50% of the ^{10}Be in ice older than 100 kyr is associated with dust grains, compared to about 5% in the Holocene and 20% in the last glacial (Baumgartner et al. 1997).

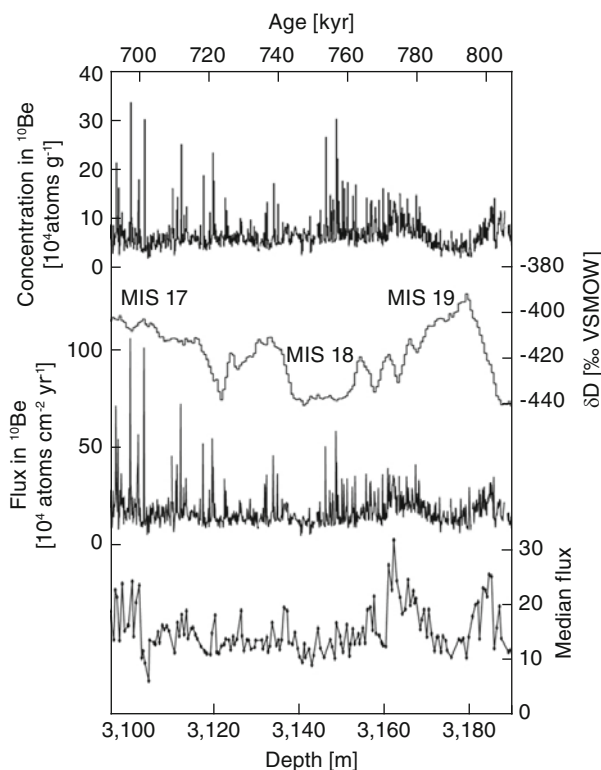
A similar phenomenon was observed in the study of the B/M reversal discussed in the previous section (Raisbeck et al. 2006). The measured ^{10}Be concentrations are shown in Fig. 21.3.2-1, together with δD , the deviation of the deuterium hydrogen ratio from the Vienna standard mean ocean water (VSMOW) for the depth interval 3,100–3,190 m (lower abscissa) corresponding to the age interval 700–800 kyr BP (upper abscissa). The marine isotope stages (MIS) reflect the numbering of the alternating warm (odd numbers) and cold (even numbers) periods of the global paleoclimate.

The ^{10}Be concentration record in the top panel is characterized by short spikes which are up to a factor of ten higher than the adjacent data points. The same is observed when the ^{10}Be fluxes (Sect. 19.4) are calculated (Fig. 21.3.2-1, lower panel). Each sample consisted of an 11-cm-long slice of ice with a diameter of only a few centimetres. Duplicate samples from the same depth partly confirmed the results, partly not, indicating a concentration effect. Calculating the median over five samples leads to a ^{10}Be flux that supports the migration hypothesis.

While ^{10}Be in very old ice seems to migrate and concentrate on dust grains, ^{36}Cl can be lost from the firm. Shortly after the development of the AMS technique, expectations were high that the $^{10}\text{Be}/^{36}\text{Cl}$ ratio would be ideally suited for dating old ice. The ratio increases exponentially with an apparent half-life of 384,000 years providing a dating interval from about 50,000 to more than one million years. However, measurements of the $^{10}\text{Be}/^{36}\text{Cl}$ ratio in young ice samples from Greenland revealed fluctuations of a factor of 2, preventing its use as a dating tool. A comparison with the $\delta^{18}\text{O}$ showed that during cold glacial times the ratio was smaller and more stable than during warmer periods.

Samples from two 5-m-deep snow pits near Vostok station (Antarctica) were analysed to investigate ^{36}Cl transport effects in the firm. Because of the small accumulation rate of about 2.1 cm of water equivalent per year, the records go back to about 1930. A comparison with the already existing ^{36}Cl record from Dye 3 in Greenland shows striking differences. The peak at Vostok is much broader and

Fig. 21.3.2-1 ^{10}Be spikes in very old ice from Dome C, Antarctica, together with δD as an indicator of the climate and the median ^{10}Be flux [figure after Raisbeck et al. (2006)]



significantly shifted towards younger ages. The maximum corresponds to 1967, almost 10 years after the peak of the ^{36}Cl producing bomb tests (Fig. 21.3.2-2). These observations can be explained on the basis that ^{36}Cl is deposited as HCl , and that metamorphic processes within the firn induced by the seasonal temperature changes cause the HCl to evaporate at a later time. The gaseous HCl then diffuses resulting in a net transport towards the surface before re-condensation occurs. This process would result in a partial loss of ^{36}Cl as well. A simple diffusion advection model explains the observations well (red solid line) (Delmas et al. 2004). This model also explains why this effect is not visible in the Dye 3 core where the accumulation rate (50 cm year^{-1}) is about 25 times larger than at Vostok.

This explanation is also consistent with the observed variable $^{10}\text{Be}/^{36}\text{Cl}$ ratio in ice cores from Greenland. Clearly, the ratio will depend on how rapidly the ^{36}Cl is out-gassing from the firn. There are two main factors controlling the out-gassing: the accumulation rate and the dust content. High accumulation reduces the time until the firn converts to ice, which varies from about 100 years at Dye 3 to 1,000 years at Vostok. The dust neutralizes the hydrochloric acid which means that ^{36}Cl is no longer volatile. The largest dust concentrations occur during glacial times, which is consistent with the observed correlation with $\delta^{18}\text{O}$.

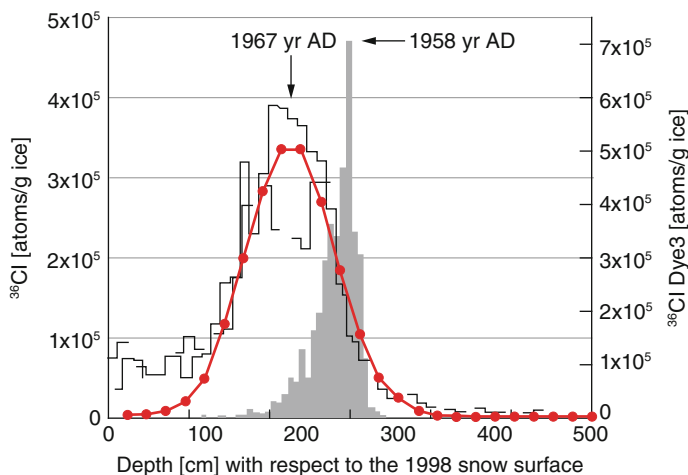


Fig. 21.3.2-2 Comparison of measured and modelled ^{36}Cl bomb profiles. The *sharp bomb pulse* peaking 1958 was measured in a shallow ice core from Dye 3, Greenland. The *broader grey curves* are compilations of two ^{36}Cl records from pits dug near Vostok station in Antarctica. The *solid red line* shows the modelled ^{36}Cl bomb pulse assuming diffuse mixing within the firn [figure after Delmas et al. (2004)]

The take home message of this discussion is that archives which appear to be almost perfect may be affected by a variety of subtle processes which may perturb the stored signal to a greater or lesser extent. It is therefore always a good strategy to confirm important results by applying independent approaches. Cosmogenic radionuclides often provide an alternative approach, as in the case of the reconstruction of the geomagnetic field intensity outlined above.

21.3.3 Transport in Soils

The behaviour of cosmogenic radionuclides deposited by wet precipitation on soil depends strongly on the geochemical properties of the respective nuclide. Although soluble nuclides such as chlorides or iodides basically follow the water cycle, particle reactive nuclides such as ^{10}Be and ^{26}Al show a complex behaviour depending on different parameters such as the mineralogical composition of the soil, grain sizes, pH, and other factors which control absorption and desorption processes and are subject to spatial and temporal changes.

The conventional way to quantify the ratio between absorbed and dissolved phases of a substance is based on the partition or distribution coefficient K_d . If we have a system consisting of solid matter (e.g. soil) and water and we add a small amount of a substance such as Beryllium, the K_d parameter tells us the percentage of Be which is absorbed by 1 g of soil divided by the percentage of Be which stays

dissolved in 1 g of water under equilibrium conditions. Obviously, the K_d coefficient depends on a large number of different parameters and in practice it can only be determined empirically. For Beryllium (^7Be and ^{10}Be) it ranges between 10^4 and 10^5 . The pH is a crucial parameter in determining K_d because it has a strong influence on the speciation of Be, that is, on the chemical compounds which Be forms in soils.

Figure 21.3.3-1 shows that the speciation of Beryllium changes with increasing pH from Be^{2+} below $p = 5$ to $\text{Be}(\text{OH})_3^-$ for $\text{pH} = 12$. If, however, humic acid is present in the soil, the situation is completely different and Be^{2+} is formed for low pH. Between $\text{pH} = 3$ and $\text{pH} = 10$ Be forms a complex with humate.

Colloids, particles with a diameter between 2 and 200 nm, play a special role in transport processes in soils. The radionuclides are absorbed onto them, and the small size of the colloids means that they remain suspended in the water and can therefore contribute considerably to transport effects. One distinguishes between inorganic colloids (clay, silicates, iron compounds) and organic colloids consisting of humic (degraded bio-molecules) and fulvic (acidic organic polymers) substances.

The penetration of meteoric ^{10}Be into soil can be considered as a sequence of sorption and desorption processes which retard the migration of ^{10}Be (in a manner similar to a chromatographic column) when rainwater containing dissolved ^{10}Be is added at the top and eluted downwards. The large K_d values mean that the downward migration of ^{10}Be is strongly retarded, leading to a strongly enhanced concentration near the surface. The concentration of adsorbed ^{10}Be increases until it reaches saturation after which the ^{10}Be is adsorbed deeper down until the whole soil column is saturated. An example of an unsaturated ^{10}Be profile measured in the Waipoa basin on the southern island of New Zealand is shown in Fig. 21.3.3-2. The ^{10}Be concentration decreases by about a factor of 5 from the surface to a depth of 1 m.

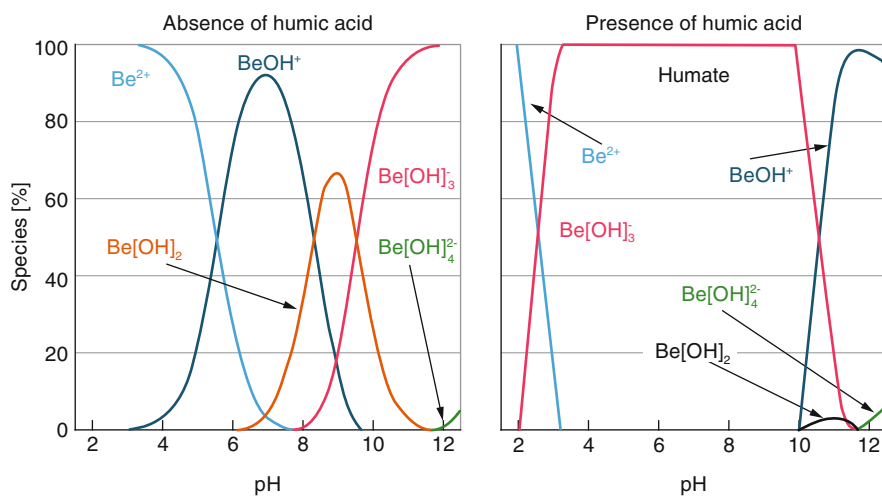
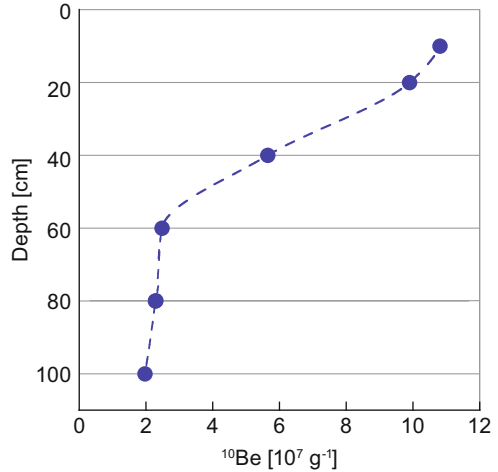


Fig. 21.3.3-1 Speciation of Beryllium as a function of pH in the absence and presence of humic acids [figure after Takahashi et al. (1999)]

Fig. 21.3.3-2 Depth profile of ^{10}Be in soils of Waipoa basin, New Zealand [figure after Reusser and Bierman (2010)]



Although the profile shown in Fig. 21.3.3-2 is far from saturation, we can use it to make a rough estimate of the K_d value. For the observed surface saturation concentration of $1.2 \times 10^8 \text{ atoms g}^{-1}$, and assuming that the concentration in rainwater is about $1 \times 10^4 \text{ g}^{-1}$, gives a K_d value of 10^4 which is in the range given at the beginning of this section.

Soil Age and Erosion Rate. Analysing the ^{10}Be budget in soils provides useful information about soil age and erosion rates. The derivation of the relevant equations is shown in Box 21.3.3.1

21.3.3.1 BOX Inventories, Age, and Erosion Rates

The total inventory I in $\text{atoms cm}^{-2} \text{ s}^{-1}$ is given by:

$$I = \int_0^D c(z) \rho dz \quad (\text{B21.3-1})$$

with $c(z)$ is the ^{10}Be concentration in atoms g^{-1} , ρ the density in g cm^{-3} , and D the thickness of the soil layer in cm.

Then the budget is:

$$\frac{dI(t)}{dt} = q(t) - \lambda I(t) - c_0(t) \rho \varepsilon(t) \quad (\text{B21.3-2})$$

with $q(t)$ the production rate in $\text{atoms cm}^{-2} \text{ y}^{-1}$, λ the decay constant in y^{-1} , ε the erosion rate in cm y^{-1} , and c_0 the surface concentration.

Assuming that $q(t)$ and $\varepsilon(t)$ are constant, Eq. (B21.3-2) can be solved:

$$I = \frac{q - \varepsilon \rho c_0}{\lambda} (1 - e^{-\lambda t}) + I_0 e^{-\lambda t} \quad (\text{B21.3-3})$$

If at the time $t = 0$ the inventory $I = 0$ we obtain

$$I = \frac{q - \varepsilon \rho c_0}{\lambda} (1 - e^{-\lambda t}) \quad (\text{B21.3-4})$$

or

$$t = -\frac{1}{\lambda} \ln \left(1 - \frac{\lambda I}{q - \varepsilon \rho c_0} \right) \quad (\text{B21.3-5})$$

If we wait long enough until steady-state conditions are reached, (B21.3-4) becomes time-independent:

$$I = \frac{q - \varepsilon \rho c_0}{\lambda} \quad (\text{B21.3-6})$$

or for the erosion rate ε :

$$\varepsilon = \frac{q - \lambda I}{\rho c_0} \quad (\text{B21.3-7})$$

If the inventory can be determined as in the case of the soil profile shown in Fig. 21.3.3-2 the erosion rate can be calculated.

Equation (B21.3-7) provides the local erosion rate for a soil layer. Often one is interested in the mean erosion rate of a river basin with locally varying depth profiles and erosion rates. In this case Eq. (B21.3-3) still holds for the whole basin with the total amount of ^{10}Be removed by erosion being balanced by the total amount deposited and removed by radioactive decay. The problem is then that we do not know the mean inventory I . However, we can approximate I by integrating the exponentially decreasing concentration:

$$c(z, t) = c_0(t) e^{-kz} \quad (\text{B21.3-8})$$

$$I(t) = \int_0^D \rho c_0(t) e^{-kz} dz \quad (\text{B21.3-9})$$

$$I(t) = c_0(t) \frac{\rho}{k} (1 - e^{-kD}) \quad (\text{B21.3-10})$$

Replacing I in (B21.3-7) by (B21.3-10) and solving for ε leads to

$$\varepsilon = \frac{q}{\rho c_0} - \frac{\lambda}{k} (1 - e^{-kD}) \quad (\text{B21.3-11})$$

where we have assumed that the concentration at the soil surface is constant and that the mean thickness of the soil layer is D .

21.3.4 *Transport in Rocks*

In theory igneous rocks such as granite are impermeable, at least on time scales comparable to the half-lives of the cosmogenic radionuclides. However, everybody who has ever visited a cave knows that in reality things are quite different. Most rocks have been fractured, and this provides the permeability necessary for fast fluid movements. Fractures, which may have been formed either in the distant past, or recently as a result of mining or other activities of mankind, are formed when the stress because of compression or tension exceeds the rock strength.

Recently, the transport of radionuclides in rocks and other unsaturated geological settings has attracted new attention in order to find safe repositories for nuclear waste. Interestingly, ^{36}Cl turned out to be one of the critical nuclides. It is produced by neutron activation of ^{35}Cl in impurities in the reactor fuel in a similar manner to the production of the ^{36}Cl bomb pulse by n-activation of sea salt (Sect. 19.3). Discharge of ^{36}Cl can then occur from a reprocessing plant or a waste disposal repository. Because of its high solubility and long half-life of 301,000 years, ^{36}Cl needs special attention in the assessment of long-term risks.

The transport of dissolved radionuclides through rocks is affected by advection in conducting fractures, matrix diffusion, and sorption onto the solid matrix. Studies in a test site in the Yucca Mountains have shown that the flow of bomb derived ^{36}Cl and ^3H is indeed related to discrete features such as fractures (Guerin 2001). Similar results from other sites confirm that these findings are ubiquitous.

Another radionuclide of interest in this context is ^{129}I which is a fission product of ^{235}U and is released in reprocessing plants either as gaseous I_2 or in water (Sect. 12.3.1).

21.3.5 *Formation of Loess Plateaus*

As discussed in Sect. 14.4 loess plateaus are formed by windblown dust originating in deserts. One of the most famous loess plateaus has been formed in China over the past approximately two million years by dust from the Gobi Desert.

Figure 21.3.5-1 shows ^{10}Be data measured in a loess profile from Lochuan together with the SECPMAP $\delta^{18}\text{O}$ stacked together from records of planktonic forams in deep-sea sediments based on a time scale tuned by orbital forcing (Imbrie et al. 1984). The time scale of the $\delta^{18}\text{O}$ was used to fine-tune the age model of the ^{10}Be record. The loess profile can be divided into a sequence of loess and palsol layers, indicating dry and cold, or wet and warm periods, respectively. For each layer the accumulation rate and the corresponding ^{10}Be flux in $\text{atoms cm}^{-2} \text{y}^{-1}$ was calculated. The results show that during cold and dry periods, when $\delta^{18}\text{O}$ is positive, the accumulation rates and ^{10}Be flux are high, compared with low values

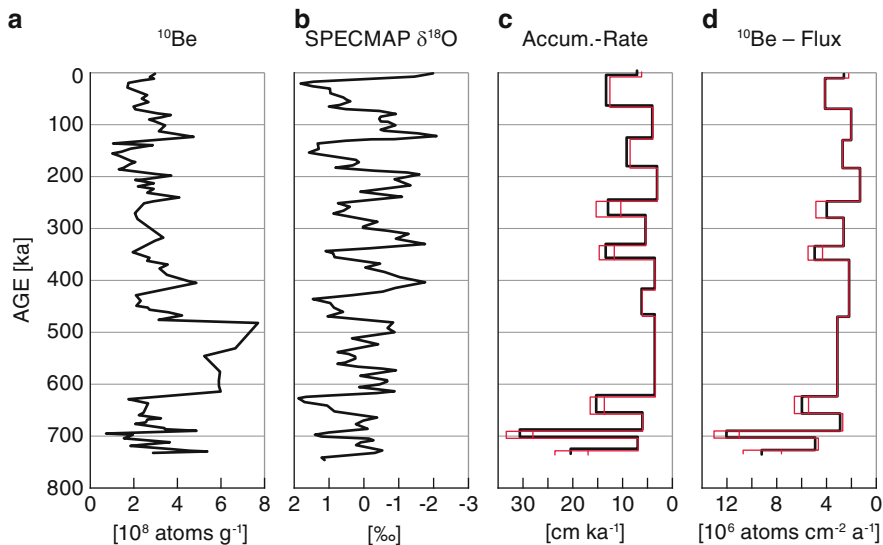


Fig. 21.3.5-1 ^{10}Be record of the Loess profile from Lochuan, China (Shen et al. 1992). The age model is tuned with the $\delta^{18}\text{O}$ SPECMAP time scale. Accumulation rate and ^{10}Be flux were calculated for the sequence of loess (cold, dry) and paleosol (warm, wet) layers

during warm and wet climatic conditions. It also shows ^{10}Be fluxes in the range of 2 to $12 \times 10^6 \text{ atoms cm}^{-2} \text{ y}^{-1}$, a factor 2 to 10 higher than the flux expected from the atmospheric production rate (Sect. 10.3).

A simple model explains the total ^{10}Be flux to the loess plateau in terms of a meteoric component representing the ^{10}Be flux originating from contemporary atmospheric ^{10}Be production, together with a dust component because of “old” ^{10}Be transported to the loess plateau attached to the dust from the Gobi Desert. This model assumes that it takes a long time for the weathering processes to produce dust grains fine enough to be uplifted by wind and carried to the loess plateau. During this time ^{10}Be from dry and wet precipitation has been adsorbed and is then transported to the loess plateau. The total ^{10}Be flux is then given by

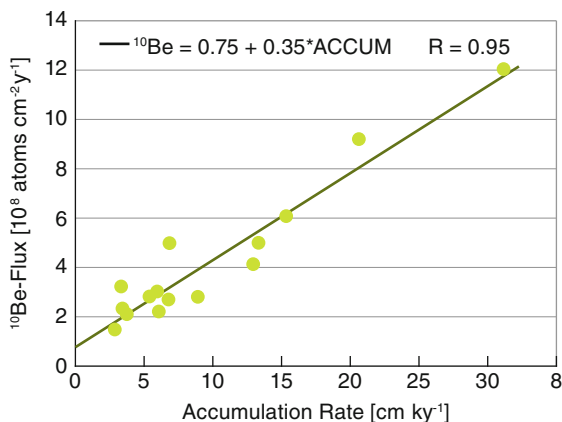
$$F_{\text{tot}} = F_{\text{m}} + F_{\text{d}} \quad (21.3.5-1)$$

where m stands for meteoric and d for dust. Inserting $F_{\text{d}} = c_{\text{d}}\rho a$ yields

$$F_{\text{tot}} = F_{\text{m}} + c_{\text{d}}\rho a \quad (21.3.5-2)$$

where c_{d} is the ^{10}Be concentration in the dust, ρ the density, and a the dust accumulation rate.

Fig. 21.3.5-2 Scatter plot between the total ^{10}Be flux and the loess accumulation rate



A scatter plot (Fig. 21.3.5-2) between total ^{10}Be flux and loess or dust accumulation rate confirms the simple model as summarized by (21.3.5-2). The intersection with the y-axis (corresponding to no dust contribution) results in a meteoric rate of $0.75 \times 10^6 \text{ cm}^{-2} \text{ y}^{-1}$ in agreement with a mean global value of about $10^6 \text{ cm}^{-2} \text{ y}^{-1}$ (Sect. 10.3.3). From the slope $c_d \rho = 0.35$, and assuming $\rho = 1.5 \text{ g cm}^{-3}$ the mean ^{10}Be dust concentration is calculated to be $c_d = 2.3 \times 10^8 \text{ g}^{-1}$. This value is consistent with a measurement on recently collected modern dust ($2.7 \times 10^8 \text{ g}^{-1}$).

21.3.6 Subduction

Plate tectonics is a fundamental process in the theory of continental drift first proposed by Alfred Wegener in the first part of the twentieth century and it plays an important role in the global carbon cycle (Sect. 13.5.3). When the relatively thin oceanic plates collide with the thicker continental plates they slide beneath them in a process called subduction. The subducted material melts, and then returns to the surface to form island arcs and volcanoes and frequently causes earthquakes. Subduction rates are slow, of the order of centimetres per year, corresponding to the continental drift rate.

Cosmogenic radionuclides such as ^{10}Be have high concentrations in recent sea sediments (of the order of 10^8 g^{-1}) and the question arises whether they would be suitable tracers for the subduction process and the subsequent formation of volcanic material.

Brown and colleagues (1982) were the first to detect ^{10}Be in basaltic samples from island-arc volcanoes. As the ^{10}Be concentrations of 2.7 to $6.9 \times 10^6 \text{ g}^{-1}$ could not be explained by in-situ production, they concluded that the source was sedimentary material from the ocean floor which had been subducted and then ejected by volcanoes.

This work was confirmed and extended by combining measurements of the ^{10}Be and ^9Be in island-arc volcanic rocks (Monaghan et al. 1988). From the linear regression between ^9Be and ^{10}Be concentrations they concluded that the $^{10}\text{Be}/^9\text{Be}$ ratio is better suited to these studies because it does not change during rock-forming processes (petrogenesis). These studies were further refined by including Boron (Morris et al. 1990) and measuring ^{10}Be depth profiles in ocean sediments along the northeast Japan and Costa Rica convergent margins (Morris et al. 2002).

All these studies clearly confirm that sedimentary material containing ^{10}Be is involved in the subduction and the magmatic processes on time scales shorter than several half-lives of ^{10}Be . However, they also show that melting, magma transport, and petrogenesis are complex processes influencing ^{10}Be and ^9Be in ways not yet fully understood.

References

- Baumgartner S, Beer J, Wagner G, Kubik PW, Suter M, Raisbeck GM, Yiou F (1997) ^{10}Be and dust. *Nucl Instrum Methods B* 123:296–301
- Brown L, Klein J, Middleton R, Sacks IS, Tera F (1982) Be-10 in island-arc volcanos and implications for subduction. *Nature* 299(5885):718–720
- Delmas RJ, Beer J, Synal HA, Muscheler R, Petit JR, Pourchet M (2004) Bomb-test Cl-36 measurements in Vostok snow (Antarctica) and the use of Cl-36 as a dating tool for deep ice cores. *Tellus B Chem Phys Meteorol* 56(5):492–498
- Guerin M (2001) Tritium and ^{36}Cl as constraints on fast fracture flow and percolation flux in the unsaturated zone at Yucca Mountain. *J Contam Hydrol* 51(3–4):257–288
- Guyodo Y, Valet J-P (1999) Global changes in intensity of the Earth's magnetic field during the past 800 kyr. *Nature* 399(may):249–252
- Imbrie J, Hays JD, Martinson DG, McIntyre A, Mix AC, Morley JJ, Pisias NG, Prell WL, Shackleton NJ (1984) The orbital theory of Pleistocene Climate: support from a revised chronology of the marine ^{18}O record. In: Berger AL et al (eds) *Milankovitch and Climate*. Reidel Publishing Company, Boston, MA, pp 269–305
- Laj C, Kissel C, Mazaud A, Channell JET, Beer J (2000) North Atlantic palaeointensity stack since 75 ka (NAPIS-75) and the duration of the Laschamp event. *Philos Trans R Soc Lond A* 358 (1768):1009–1025
- Masarik J, Beer J (2009) An updated simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res-Atmos* 114:D11103
- McCracken KG (2004) Geomagnetic and atmospheric effects upon the cosmogenic Be-10 observed in polar ice. *J Geophys Res-Space Phys* 109:A04101
- McElhinney MW, McFadden PL (2000) *Paleomagnetism: continents and oceans*. Academic, San Diego, CA
- Monaghan MC, Klein J, Measures CI (1988) The origin of ^{10}Be in island-arc volcanic rocks. *Earth Planet Sci Lett* 89:288–298
- Morris JD, Leeman WP, Tera F (1990) The subducted component in island arc lavas: constraints from Be isotopes and B-Be systematics. *Nature* 344:31–37
- Morris J, Valentine R, Harrison T (2002) ^{10}Be imaging of sediment accretion and subduction along the northeast Japan and Costa Rica convergent margins. *Geology* 30(1):59–62
- Muscheler R, Beer J, Kubik PW, Synal H-A (2005) Geomagnetic field intensity during the last 60,000 years based on ^{10}Be & ^{36}Cl from the Summit ice cores and ^{14}C . *Quat Sci Rev* 24:1849–1860

- Raisbeck GM, Yiou F, Cattani O, Jouzel J (2006) Be-10 evidence for the Matuyama-Brunhes geomagnetic reversal in the EPICA Dome C ice core. *Nature* 444(7115):82–84
- Reusser LJ, Bierman PR (2010) Using meteoric ^{10}Be to track fluvial sand through the Waipaoa River basin, New Zealand. *Geology* 38(1):47–50
- Shen CD, Beer J, Liu TS, Oeschger H, Bonani G, Suter M, Wolfli W (1992) Be-10 in Chinese loess. *Earth Planet Sci Lett* 109(1–2):169–177
- Singer BS, Pringle MS (1996) Age and duration of the Matuyama-Brunhes geomagnetic polarity reversal from incremental heating analyses of lavas. *Earth Planet Sci Lett* 139(1–2):47–61
- Singer BS, Guillou H, Jicha BR, Laj C, Kissel C, Beard BL, Johnson CM (2009) Ar-40/Ar-39, K-Ar and Th-230-U-238 dating of the Laschamp excursion: a radioisotopic tie-point for ice core and climate chronologies. *Earth Planet Sci Lett* 286(1–2):80–88
- Takahashi Y, Minai Y, Ambe S, Makide Y, Ambe F (1999) Comparison of adsorption behavior of multiple inorganic ions on kaolinite and silica in the presence of humic acid using the multitracer technique. *Geochim Cosmochim Acta* 63(6):815–836

Chapter 22

Biosphere

22.1 Introduction

The term “biosphere” was introduced by Eduard Suess at the end of the nineteenth century to describe “The place where life dwells”. It embraces the sum of all ecosystems and deals with the effects that living beings have on the atmosphere, the upper part of the geosphere, and the hydrosphere. It is impressive that living systems have inhabited almost every part of Earth, from the cold polar caps, to the hot tropics, and from the deepest part of the oceans, up to altitudes of 40 km in the atmosphere.

Carbon-based life developed on Earth about four billion years ago. Some time later, algae developed that used photosynthesis to turn CO_2 into organic matter and O_2 . As a result our atmosphere consists now of about 21% of oxygen. Bacteria that decompose organic matter produce methane in the absence of oxygen. Many more trace gases which are very important for the radiation balance and the climate on Earth are related to life on Earth. It is no wonder that James Lovelock concluded that the best way to detect life on a planet is to analyse its atmosphere (Lovelock 1965).

Vegetation is a major component of the biosphere and has an important impact on the Earth in general. It affects the albedo, the fraction of solar radiation reflected back to space, the formation of soils, the erosion of soils and rocks, the hydrological cycle, and all biogeochemical cycles in general. In recent times human activities have played an increasingly important role which led to the definition of a new sphere, the “anthroposphere”. Quantification of these and many other matters are important in order to reach a better understanding of the biosphere of which we are a part.

With regard to cosmogenic radionuclides the biosphere is a relatively new field of endeavour, but at the same time one of the most promising. In addition there are a growing number of applications based on artificially produced radionuclides. Living organisms are vulnerable to radioactivity and therefore it is important that any radionuclides used as tracers in the study of processes within the biosphere

should have relatively low activity. It is obvious that ^{14}C plays a central role in the biosphere because it is by definition part of organic material, and therefore it is an ideal tracer. To this end artificially produced ^{14}C can be built into complex organic molecules and used to study specific biochemical processes.

All these interesting applications are beyond the scope of this book but should be borne in mind as areas of potential importance. In the following we restrict our attention to the application of the mostly naturally produced radionuclides ^{14}C , ^{36}Cl , and ^{129}I in the investigation of several aspects of the biosphere.

22.2 Radiocarbon Applications

As discussed in the dating section (Chap. 23) a living plant exchanges CO_2 with the atmosphere until it dies and the radiocarbon clock then starts to tick. Animals feeding on plants and carnivores eating other animals will all be synchronized and therefore datable with radiocarbon. There are exceptions however that must be borne in mind. For example, consider an Inuit in Greenland who lives exclusively on fish and other seafood for some months at a time. If we were to radiocarbon date his hair grown during this period we would probably be surprised to obtain an age of about 400 years. How is this possible? The simple explanation is that the carbon in his body is not coming from the atmosphere but from the surface of the ocean. As we have discussed the residence time of ^{14}C in the ocean is more than a thousand years leading to a smaller $^{14}\text{C}/^{12}\text{C}$ ratio than in the atmosphere. In the surface water the ratio is 5% lower than in the atmosphere corresponding to an age of about 400 years. This value is determined mainly by the strength of the thermohaline circulation (THC) and represents the present situation. There is a clear indication that during glacial times the abrupt surges of melt water from the continents into the ocean reduced and changed the THC. These so-called Dansgaard-Oeschger events caused temperature changes of up to 15°C within 1–2 decades in Greenland and $4\text{--}6^\circ\text{C}$ in the Tropics (Broecker 1996).

Understanding the THC is crucial because it acts like a conveyor belt transporting heat from low latitudes to the polar regions, and this has a major effect upon climate, and in turn on the biosphere. Understandably then, attempts have been made to detect changes in the THC by using the marine biosphere itself. Foraminifera or forams are organisms living in the oceans: they are typically smaller than one millimetre in diameter and form tiny carbonate shells which settle down into the sea floor sediments after the death of the forams. In fact, limestone is composed to a large extent of forams. These shells provide measurements of $\delta^{18}\text{O}$, the famous paleothermometer, and also of the $^{14}\text{C}/^{12}\text{C}$ ratio of the water they lived in.

Most forams live near the bottom of the sea (called benthic forams) where the $^{14}\text{C}/^{12}\text{C}$ ratio is smaller than that experienced by the planktonic forams which live in the surface water. Because they have different shapes they can be separated in samples of the sea-floor sediment and radiocarbon dated separately. The age

difference between the planktonic and benthic forams then provides a direct measure of the water overturning time. Figure 22.2-1 shows an example from the Bermuda rise (Roberts et al. 2010). Although there are some problems (for example due to bioturbation in the sediment), there is a clear decrease in age difference from the last glacial maximum at 20,000 BP towards the Holocene starting at 11,600 BP. A corresponding warming is visible in the $\delta^{18}\text{O}$ record from the GRIP ice core (Greenland). The warming trend was interrupted by an abrupt return to glacial conditions at about 12,900 BP. This cold event is called the Younger Dryas (YD) and lasted about 1,300 years. The Younger Dryas is generally thought to be the result of a sudden outbreak of Lake Agassiz (a gigantic lake which formed in the centre of North America during the end of the last glaciation) into the North Atlantic that reduced the THC considerably. The Younger Dryas is clearly visible in both the $\delta^{18}\text{O}$ and age difference graphs in Fig. 22.2-1.

We now consider the extent to which cosmogenic ^{14}C contributes to the intrinsic radioactivity of the animals living in the biosphere. Thus it is well known that radioactivity is harmful to living organisms because the energy released in a disintegration can destroy cells or the DNA within the cell. To be specific, we now compute the average ^{14}C disintegration rate in a body of a 70 kg human. The result depends to some extent on where the person lives and what it eats. Approximately 70% of a person's weight is water and 20% carbon. This leads to about 15 kg of carbon or 7.5×10^{26} ^{12}C atoms per person. As the $^{14}\text{C}/^{12}\text{C}$ ratio in today's atmosphere is 1.2×10^{-12} , the number of ^{14}C atoms is 9×10^{14} . This leads to an

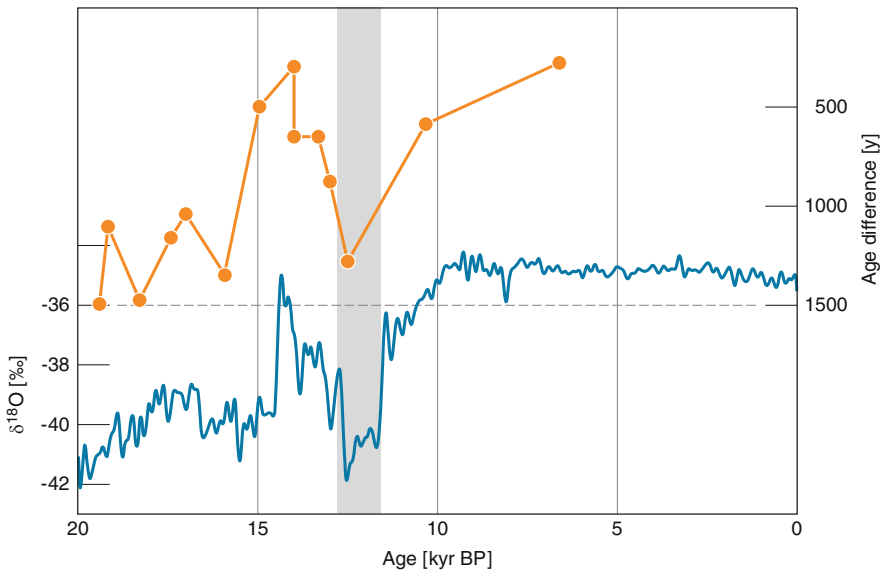


Fig. 22.2-1 Age differences between benthic (sea floor) and planktonic (sea surface) forams from the Bermuda rise (orange line) [after (Roberts et al. 2010)] with $\delta^{18}\text{O}$ from the GRIP ice core in Greenland reflecting the temperature (blue line)

activity of 3,500 disintegrations per second. In other words, 3,500 ^{14}C atoms decay in our body every second. Furthermore, additional radionuclides such as ^{40}K more than double the ^{14}C activity, to which is to be added the mesons and electrons from the nucleonic cascades initiated by the cosmic rays arriving from outer space.

As discussed in Chaps. 10 and 23 the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio varies as a consequence of changes in the cosmic ray intensity due to solar and geomagnetic activity, and nuclear bomb tests. The ratio will also change if CO_2 enters the atmosphere which has a different $^{14}\text{C}/^{12}\text{C}$ ratio. This occurs naturally in volcanic eruptions which emit “dead” CO_2 (i.e. CO_2 formed from very old carbon in which any ^{14}C has decayed). While this natural process is small and has been going on continuously in the past, anthropogenic CO_2 (also largely devoid of ^{14}C) has been added in increasing quantities since the beginning of the nineteenth century. Thus industrialization has led to an ever-growing consumption of fossil fuel such as coal, oil, and natural gas. All of these were formed from plants and other organic material buried underground for periods of more than a hundred million years. Except for a very small amount of ^{14}C produced in situ they only contain ^{12}C and ^{13}C and therefore dilute the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio (Fig. 22.2-2). This is called the “Suess-effect” after its discoverer, Hans Suess.

In Fig. 22.2-2 the peak between 1650 and 1720 is the consequence of the Maunder Minimum period of very low solar activity. The rapid decrease that started after 1800 is mainly due to the combustion of fossil fuel which caused an increase in the atmospheric CO_2 concentration and a corresponding decrease of the $^{14}\text{C}/^{12}\text{C}$ ratio. The minor peaks near 1800 and 1900 are because of periods of low solar activity. The steep slope between 1900 and 1950 is partly because of increasing solar activity, but the majority is of anthropogenic origin. Finally the decreasing trend of the global dipole moment reduces the $\Delta^{14}\text{C}$ slope by a few permil.

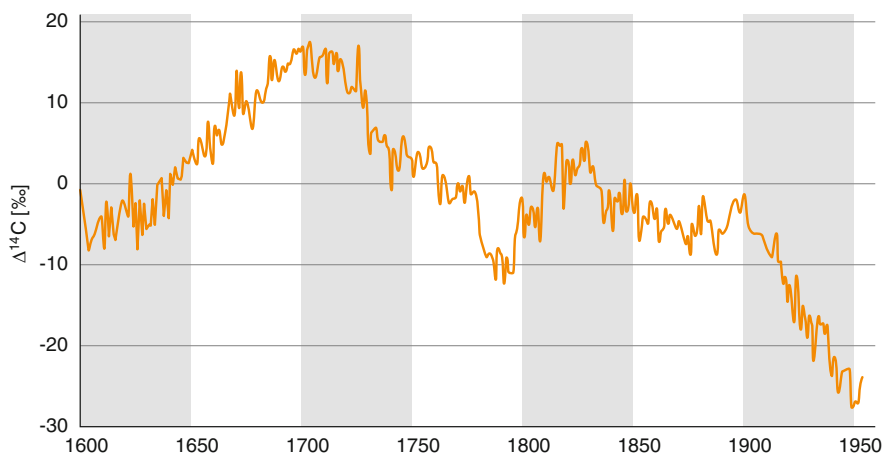


Fig. 22.2-2 The drop in $\Delta^{14}\text{C}$ after 1800 is to a large extent the result of the combustion of fossil fuel which reduces the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio (Suess effect)

22.3 Chlorine-36 in Ecosystems

^{36}Cl is highly soluble which makes it also highly mobile in soil. Chlorine is a biologically important element with a large bioavailability which means that a biological system incorporates it easily and efficiently. Its long half-life of 301,000 years, its large production peak during the nuclear bomb tests in the 1950s, and its presence in nuclear waste make it one of the key nuclides in the field of radioecology. So far, little is known about the behaviour of ^{36}Cl in the biosphere and that is the reason why the biosphere is not included in the model biogeochemical cycle of chlorine depicted in Fig. 13.5.4-1. It is known that more than 400 organic compounds containing chlorine are produced in the oceans by seaweeds and other algae. Many of them such as methyl chloride and chloroform are volatile, resulting in a net flux into the atmosphere.

An important question is how ^{36}Cl gets into the food chain. As an illustration of the complexity of the ^{36}Cl uptake, Fig. 22.3-1 shows the concentration of ^{36}Cl that occurs in a number of plants. This experiment was carried out in the laboratory (Colle et al. 2002) using three different crops (radishes, lettuces, and beans) planted in three different types of soil (calcareous, acidic, and organic). Twenty millilitres of water containing ^{36}Cl in the form of NaCl was added to the soil and well mixed. Then the vegetables were planted. After 2 to 3 weeks the plants were dried at 60°C and analysed. The large ratios of up to almost 400 show the potential of biological system to concentrate radionuclides. The range of the concentration factor is very variable and depends on the soil type as well as on the plant. The ^{36}Cl uptake is largest for calcareous soils and smallest for organic soils. The distribution within the plants shows a clear preference for leaves and stems compared to roots and fruits.

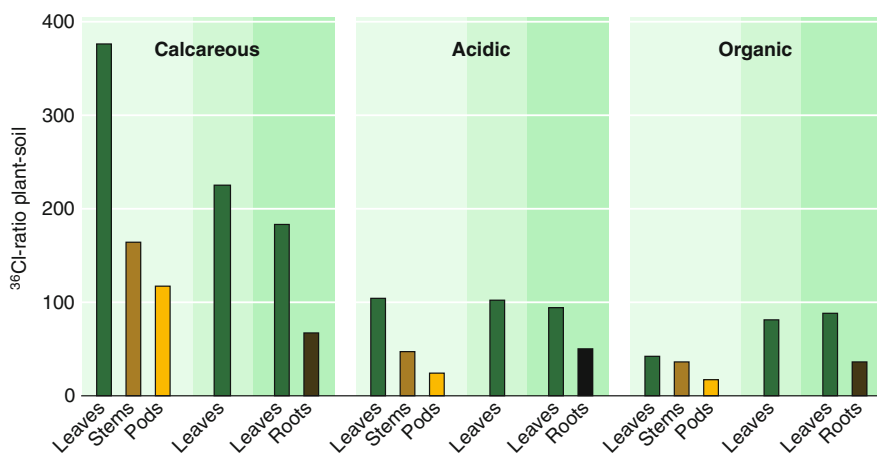


Fig. 22.3-1 Ratio of ^{36}Cl activity in Bq kg^{-1} between dry plant material and soil in which the plants were grown

22.4 Iodine-129

Iodine plays an important role in the body in the synthesis of thyroid hormones by the thyroid gland. These hormones are essential for (1) growth, (2) operation of the nervous system, and (3) metabolism. Large quantities of iodine are dangerous, causing disturbed heartbeats and loss of weight. ^{131}I with a half-life of 8.02 days was one of the radionuclides with the largest short-term health risk released during the Chernobyl and Fukushima accidents. It appeared in the food chain (e.g. through milk of cows feeding on grass) and caused an increased rate of thyroid cancer (Bleuer et al. 1997). On the other hand, ^{131}I is used in radiation therapy to treat thyrotoxicosis and thyroid cancer.

^{129}I is produced naturally by cosmic ray-induced spallation of xenon in the atmosphere (Chap. 10) and spontaneous fission of uranium in the lithosphere. Both processes lead to an estimated global inventory of 50 tons. However, only about 250 kg is exchanged between the atmosphere, hydrosphere, and biosphere (see Sect. 13).

During the past decades large amounts of ^{129}I have been released by nuclear waste-reprocessing plants. The two main sources are the European plants at La Hague (France) and Sellafield (UK) which contribute 70% and 30%, respectively. Up to 2000 the total emissions amounted to about 3,500 kg, 95% in the liquid and 5% in gaseous form (Michel et al. 2005). This increased the natural inventory by more than an order of magnitude, resulting in disequilibrium between the different compartments of the iodine cycle. Early equilibrium $^{129}\text{I}/^{127}\text{I}$ ratios were estimated to be approximately 5×10^{-13} in the marine hydrosphere and only slightly higher in the continental atmosphere, the terrestrial biosphere, and in soils (Michel et al. 2005). Today the $^{129}\text{I}/^{127}\text{I}$ ratios are much higher: North Sea: 1.5×10^{-6} ; organic and inorganic soil particles in northern Germany: 1.8×10^{-7} . The reprocessing plants Sellafield in UK and La Hague in France represent point sources of ^{129}I in the Atlantic and measurements in archived seaweeds provide the means to study the water currents with an unprecedented temporal and spatial resolution (Yiou et al. 2002).

22.5 Aluminium-26

Aluminium is the most abundant metal on Earth and can be found in soil, water and air. It has specific chemical and physical properties that make it suitable for many different applications. It became very popular in consumer products (foil, cooking ware), in food additives and drugs (antacids), and in the treatment of drinking water. There has been some concern regarding the potential health effects because of a growing intake of aluminium in humans. ^{26}Al is a powerful tracer that permits study of such effects and of the biokinetics of ^{26}Al in general. The natural ^{26}Al background is negligible and consequently a single dose of ^{26}Al can be low enough

neither to disturb the natural budget nor to cause saturation effects. Last and not least the long half-life of 730,000 years guarantees a negligible radiation dose for the patient. Several studies have addressed the absorption, distribution, speciation, and excretion of different aluminium compounds in animals and humans (Jouhanneau et al. 1997).

References

- Bleuer JP, Averkin YI, Abelin T (1997) Chernobyl-related thyroid cancer: what evidence for rose of short-lived iodines? *Environ Health Perspect* 105:1483–1486
- Broecker W (1996) Glacial climate in the tropics. *Nature* 272:1902–1904
- Colle C, Mauger S, Massiani C, Kashparov VA, Grasset G (2002) Behaviour of chlorine 36 in cultivated terrestrial ecosystems. *Radioprotection* 37(C1):491–496
- Jouhanneau P, Raisbeck GM, Yiou F, Lacour B, Banide H, Druke TB (1997) Gastrointestinal absorption, tissue retention, and urinary excretion of dietary aluminum in rats determined by using Al-26. *Clin Chem* 43(6):1023–1028
- Lovelock JE (1965) A physical basis for life detection experiments. *Nature* 207(4997):568–570
- Michel R, Klipsch K, Ernst T, Gorny M, Jakob D, Vahlbruch J, Synal H-A, Schnabel C (2005) Long-lived radionuclides in the environment: on the radioecology of Iodine-129. *Radioprotection* 40(suppl 1):S269–S276
- Roberts NL, Piotrowski AM, McManus JF, Keigwin LD (2010) Synchronous deglacial overturning and water mass source changes. *Science* 327(5961):75–78
- Yiou F, Raisbeck GM, Christensen GC, Holm E (2002) I-129/I-127, I-129/Cs-137 and I-129/Tc-99 in the Norwegian coastal current from 1980 to 1998. *J Environ Radioact* 60(1–2):61–71

Chapter 23

Dating

23.1 Introduction

So far we have discussed the aspects of the cosmogenic radionuclides which are mainly related to their “colour”, a terminology that refers to their slightly different masses which enables us to follow their pathways and to trace a large variety of environmental processes. A second more specific property is that they also carry a “clock” which means that they can provide time information. Most of the cosmogenic isotopes which are found in the environment are unstable and decay. The daughter isotope may decay again as another step in the “decay chain” until a stable combination of proton and neutrons in the nucleus is reached and the decay chain comes to an end. The decay of a given radionuclide is described by one single parameter, the decay constant λ , which is the probability that an individual atom is going to decay within a certain time interval.

The activity of a sample is then given by the number of atoms times the decay constant:

$$A = N\lambda \quad (23.1-1)$$

The decay constant of ^{14}C is $1/(8267 \text{ years})$ which means that if we have one million ^{14}C atoms about 120 of them will decay in 1 year. We cannot say anything about whether a particular atom will decay; all we know is how a large number of them will behave. The more atoms we have, the more precise our prediction becomes about how many will decay within a certain time period. It is like playing with a dice. We have no idea which number we will get the next time, but we know very precisely that if we throw the dice six million times we will get a six 1 million times.

Using the decay constant λ the rate of change in the number of radionuclides, N , is given by:

$$dN = -\lambda N dt \quad (23.1-2)$$

which leads to the well-known “law of radioactive decay”:

$$N(t) = N_0 e^{-\lambda t} \quad (23.1-3)$$

where N_0 is the initial number of atoms and $N(t)$ is the number after time t . In other words, the number of radioactive atoms decreases exponentially with time.

Taking the natural logarithm of both sides of (23.1-3), and rearranging yields the equation for elapsed time, t , that is the basis of all radiometric dating

$$t = \frac{1}{\lambda} \ln \left(\frac{N_0}{N(t)} \right) \quad (23.1-4)$$

The time for half of the atoms to decay is called the half-life, $T_{1/2}$ the half-life. It is obtained by inserting $(N_0/N(t)) = 2$ into (23.1-4), thus

$$T_{1/2} = \frac{\ln(2)}{\lambda} = \frac{0.693}{\lambda} \quad (23.1-5)$$

The parameter λ in the above equations is a physical constant which does not change with time and is usually available from reference material (however, see below how to determine the half-life) and the number of atoms $N(t)$, or the activity $\lambda N(t)$, at time t is what we measure when dating a sample. The critical and often unknown parameter is N_0 , the number of atoms or the activity at time zero. Before we address the question of how to determine N_0 in detail we discuss dating in a more general way.

Any change in the number of atoms which is monotonic and follows a well-defined law can be used for dating. In the case of cosmogenic radionuclides there are two sources of change, the production and the decay which can be expressed mathematically as follows:

$$dN/dt = P(t) - \lambda N \quad (23.1-6)$$

where $P(t)$ is the production rate as a function of time, leading to the general solution:

$$N(t) = e^{-\lambda t} \left(\int_0^t P(t) e^{\lambda t} dt + C_0 \right) \quad (23.1-7)$$

If $P(t)$ is constant and equal to P_0 , the solution of this differential equation is:

$$N(t) = \frac{P_0}{\lambda} (1 - e^{-\lambda t}) + C_0 e^{-\lambda t} \quad (23.1-8)$$

where C_0 is the number of atoms that were present when the production started.

As an example, we now consider a closed system which is initially free of any radionuclides ($C_0 = 0$ at $t = 0$). Then at time $t = 50$ the production changes from 0 to 1 until $t = 200$, when it is stopped again (Fig. 23.1-1). Setting λ to 0.05 (corresponding

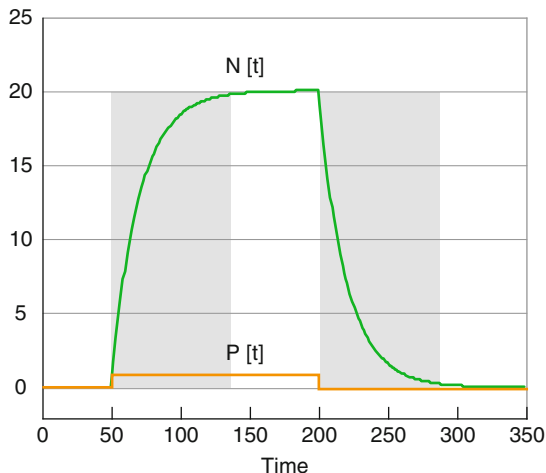


Fig. 23.1-1 Illustrating the principles of build-up and decay dating according to (23.1-8). The production rate is switched from zero to one during the time 50–200. This leads to an exponential increase followed by an exponential decrease after turning off the production. The rates of rise and decay of the curve, and the equilibrium value, depends on the decay constant (0.05 in this example). For both the build-up and the decay, a measurement of $N(t)$ can be used for dating for a period which is generally less than 10 half-lives (*shaded areas*)

to a half-life of $0.6931/\lambda = 13.862$, $N(t)$ starts growing exponentially at time $t = 50$ and asymptotically approaches the value of 20 ($=1/0.05$) when production and decay are in equilibrium. After turning the production off at $t = 200$, $N(t)$ decreases exponentially and asymptotically approaches zero again. In the build-up and the decay portions of the curve, each value of $N(t)$ corresponds to a unique value of time t . In practice, the decreasing slope of the curve at late times means that a small error in $N(t)$ leads to a relatively large error in the estimate of t . As a general rule, build-up and decay dating can only be used for an age interval of about 10 half-lives, in which 99.9% of the equilibrium value has been reached, or 99.9% of the atoms have decayed, respectively. This range is indicated in Fig. 23.1-1 by the shaded areas.

23.2 Absolute Dating

In this section we discuss the practical details of how the law of radioactivity is used to get absolute ages, and in particular how it is applied when the original number of atoms, N_0 , is not known. In this we will distinguish two different forms of dating: (1) “decay dating” based on the well-defined decrease in the number of atoms due to radioactive decay, and (2) “build-up dating” which determines the number of atoms produced “in situ” by the cosmic radiation. Finally we discuss how calibration curves and the ratios of different radioisotopes can be used to overcome the problem of not knowing the initial value of N_0 in (23.1-4).

We use the example of ^{14}C in the following discussion of decay dating because it is the most important radionuclide as far as dating is concerned, and because it illustrates the issues that apply to other radioisotopes. Based on (23.1-4), it was initially assumed that N_0 did not change much with time. In retrospect, this approach was dangerous because we know (Chap. 10) that the production rate of cosmogenic radionuclides is changing on all time scales. System effects such as transport, exchange between reservoirs and so on make things even worse (for some radionuclides). Experience has shown, however, that in the case of ^{14}C the reservoirs (ocean, atmosphere and biosphere) are large enough to smooth out short-term production changes to a reasonable degree.

The obvious way to test the assumption of a constant N_0 was to determine the ages of samples of known age. Historically the first tests were done on archeologically dated material such as mummies from Egypt. Within the experimental uncertainties, the majority of the radiocarbon ages agreed with the historical ages. Motivated by this success, people started to look into the question of the stability of N_0 in greater detail. To this end they used tree rings which provide annual resolution, and can be dated to within 1 year by counting back from the present (see Sect. 14.4). Quite soon it became clear that N_0 was only stable within about 10%, and that it showed a long-term trend with superimposed short-term fluctuations.

This brings us to the second approach to absolute dating, that of calibrating for known changes in N_0 . Tree rings, etc., were used to determine N_0 as a function of time and this was then used to correct the estimates based on a constant N_0 . This concept of first determining a “radiocarbon age”, and then correcting it with a calibration curve to a “calibrated age” is still used in radiocarbon dating. The graph of N_0 versus time is called the “calibration curve” (Fig. 23.2.1-1) wherein the changes in N_0 are expressed as a deviation of the $^{14}\text{C}/^{12}\text{C}$ ratio from a standard value in permil (see below).

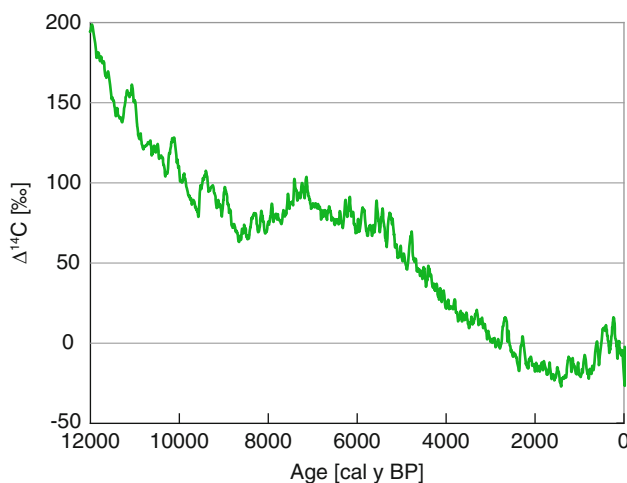


Fig. 23.2.1-1 $\Delta^{14}\text{C}$ for the past 12,000 years, after INTCAL04 (Reimer et al. 2004)

23.2.1 Principle of Radiocarbon Dating

As outlined above, radiocarbon dating is the oldest and still the most important radiometric dating technique. Every year probably several tens of thousands of samples are analyzed for ^{14}C . The method was invented by the Nobel Laureate Williard Libby in the 1940s. He put elemental carbon from various historically dated objects such as wood and Egyptian mummies into a proportional gas counter and demonstrated that the ^{14}C activity decreased with age according to the law of radioactive decay (Eq. 23.1-3). In the following years, the detection technique was considerably improved using proportional and liquid scintillation counting techniques. The development of AMS in the 1980s (Sect. 15.3) did not improve the precision, but reduced the sample size by 3–5 orders of magnitude, thus opening up completely new fields of applications and considerably increased the sample throughput.

Why is radiocarbon dating so important? There are many reasons: (1) carbon is present in comparatively large quantities in all kinds of organic matter, which opens a wide field of applications. (2) The half-life of 5,730 years provides a dating range of about 40,000 years, a period that covers the most important part of the history of mankind. (3) The production rate of ^{14}C is $\sim 2 \text{ atoms cm}^{-2} \text{ s}^{-1}$, the largest of all cosmogenic radionuclides. (4) It is globally well mixed and short-term production changes are strongly dampened by the carbon cycle (Sect. 13.5.3). As a result, the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio is spatially homogeneous, and varies with time in a well-determined manner.

23.2.1.1 BOX Isotopic Fractionation

The chemical properties of an atom are basically determined by the configuration of its electrons. Each atom of a given element has the same number of electrons and an equal number of protons in its nucleus. Frequently, however, there are several different “isotopes” of a given element, with different numbers of neutrons in the nucleus. For example, carbon can contain six, seven, or eight neutrons together with six protons leading to ^{12}C , ^{13}C , and ^{14}C with relative abundances of 98.89, 1.11, and $1.2 \times 10^{-10}\%$, respectively. Although the electron configurations are identical for these three isotopes their mass differences cause subtle effects leading to an enrichment of one isotope relative to another in chemical and physical processes. This enrichment is called isotopic fractionation. It is expressed in the so-called δ notation where

$$\delta^{13}\text{C} = \left[\frac{(^{13}\text{C}/^{12}\text{C})_{\text{sample}} - (^{13}\text{C}/^{12}\text{C})_{\text{standard}}}{(^{13}\text{C}/^{12}\text{C})_{\text{standard}}} \right] \cdot 1,000 \quad (\text{B23.2.1.1-1})$$

(continued)

is the relative deviation of the $^{13}\text{C}/^{12}\text{C}$ ratio from a standard in permil. In the case of photosynthesis, the lighter carbon isotope ^{12}C is preferentially incorporated into organic matter. The fractionation can be measured with a mass spectrometer and is found to be different for different plants, with $\delta^{13}\text{C}$ ranging from -6 to -34‰ . For ^{14}C we assume that the fractionation is twice as large as that for ^{13}C because it is mass dependent. Therefore, for each radiocarbon sample $\delta^{13}\text{C}$ is measured as well as $^{14}\text{C}/^{12}\text{C}$, and the result is used to correct the ^{14}C ratio for the effects of fractionation.

For a detailed discussion of all the conventions used in radiocarbon dating, we refer to Stuiver and Polach (1977).

As long as an animal or a plant on land is alive, it is in exchange with the atmosphere and the $^{14}\text{C}/^{12}\text{C}$ ratio in its new growth is almost identical to the atmospheric ratio. The reason that it is not completely identical is due to isotopic fractionation (see Box 23.2.1.1). As soon as it dies, the connection with the atmosphere is broken and the decay clock starts ticking. From (23.1-3) we can get:

$$R(t) = R_0 e^{-\lambda t} \quad (23.2.1-1)$$

where $R = ^{14}\text{C}/^{12}\text{C}$ and R_0 is the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio at the time the exchange with the atmosphere ceased. λ is $\ln(2)/5,730 = 1/(8,267 \text{ years})$. In a manner analogous to (23.1-4), the age can then be calculated:

$$t = \frac{1}{\lambda} \ln \frac{R_0}{R} \quad (23.2.1-2)$$

Under ideal conditions R_0 might be constant and it could then be calculated from this equation. However, as discussed above, and from Chap. 10, we know that the cosmic ray intensity, and consequently the concentration of ^{14}C and thence R_0 are functions of time. By dating tree rings of known age, a continuous record of R_0 has been obtained that is used to improve dating accuracy, as outlined below.

Presently the $^{14}\text{C}/^{12}\text{C}$ ratio, R , is about 1.2×10^{-12} . To avoid using such a small ratio conventionally the deviations of R from a standard value, R_s , is considered. By definition then the relative deviation of R in permil, $\Delta^{14}\text{C}$, is given by.

$$\Delta^{14}\text{C} = \frac{R - R_s}{R_s} \cdot 1,000 = \left(\frac{R}{R_s} - 1 \right) \cdot 1,000 \quad (23.2.1-3)$$

Figure 23.2.1-1 presents a compilation of $\Delta^{14}\text{C}$ data covering the past 12,000 years (Reimer et al. 2004). By convention, “time before present” (BP) is relative to the year 1950. The most striking feature of this curve is that R_0 shows a clear

long-term trend with generally increasing $^{14}\text{C}/^{12}\text{C}$ ratios going back in time. Superimposed on this long-term trend are the so-called “Suess” wiggles, short-term positive deviations. As discussed in Sect. 13.5.3 these long- and short-term features are the combined effect of solar and geomagnetic modulation of the cosmogenic radionuclide production rate and of the memory effect of the carbon system. It is important to note that this calibration curve has been established from tree rings and therefore reflects the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio only. As we have seen in Sect. 20.3 the $^{14}\text{C}/^{12}\text{C}$ ratio in lakes and the ocean is lower as a consequence of dissolution of old carbonate and long carbon residence times. Therefore any organism feeding on this carbon has a correspondingly older radiocarbon age.

At this point we are interested on the effects of these fluctuations upon the date obtained for a given value of $R(t)$ if we were to use $R_0 = R_S$ in (23.2.1-2). For the moment, taking $R(t)$ as a constant, the derivative of (23.2.1-1) shows that:

$$\frac{dR_0}{dt} = \lambda R e^{\lambda t} = \lambda R_0 \quad (23.2.1-4)$$

$$\frac{dR_0}{R_0} = \lambda dt = \frac{1}{8,267} dt \quad (23.2.1-5)$$

so the error in the inferred age, Δt , is given by

$$\Delta t \approx 8.3 \cdot \Delta^{14}\text{C} \quad (23.2.1-6)$$

That means a difference of 1 permil in $\Delta^{14}\text{C}$ results in a dating error of about 8.3 years. Figure 23.2.1-1 shows that $\Delta^{14}\text{C}$ has been within the range $\pm 20\%$ for the last 4,000 years, so radiocarbon dating with a constant R_0 in that interval leads to deviations from the true age of less than 170 years. However, the value of $\Delta^{14}\text{C} = 200\%$ for 12000 BP implies an error of 1,700 years. Using the values of R_0 in Fig. 23.2.1-1, however, these errors can be reduced, as outlined below.

By convention, radiocarbon dating is a two-step process. First, the “radiocarbon age” is determined based on the assumption of a constant R_0 and by applying a correction for the fractionation based on $\delta^{13}\text{C}$. Then the “calibrated age” (or “cal” age) is derived using the calibration curve (Fig. 23.2.1-1). There are very good reasons for using this two-step process. While the first step needs only be done once, the calibration curve is subject to continuous improvements. Having the radiocarbon age allows the user to recalibrate the data at a later date using the most recent calibration curve.

There is a second reason for this two-step procedure. When Libby invented radiocarbon dating the half-life of ^{14}C was thought to be 5,568 years instead of the presently accepted value of 5,730 years. In view of the uncertainty whether there would be a succession of revisions of the half-life, and to avoid confusion, the radiocarbon community decided to consistently use the old half-life of 5,568 years for calculating radiocarbon ages. Using the wrong half-life may look a bit funny,

however, since the radiocarbon age has to be calibrated anyway, it was decided to do so to guarantee that all radiocarbon ages are reported in a consistent manner, and that they can be compared directly.

To calibrate a radiocarbon age the calibration curve (Fig. 23.2.1-1) is used in a different form in which the radiocarbon age is plotted versus calibrated or calendar age. The relationship between $\Delta^{14}\text{C}$, the radiocarbon age, and the calibrated age can be derived as follows:

According to (23.2.1-3) $\Delta^{14}\text{C}$ can be expressed by the relative deviation of the measured $^{14}\text{C}/^{12}\text{C}$ ratio from the $^{14}\text{C}/^{12}\text{C}$ ratio of a standard. Now we use the same terminology for the radiocarbon ratio R_C and the absolute or dendro-age ratio R_D .

$$\Delta^{14}\text{C} = \left(\frac{R_C}{R_R} - 1 \right) \cdot 1,000 = \left(\frac{R(t)e^{\lambda_C t_C}}{R(t)e^{\lambda_R t_R}} - 1 \right) \cdot 1,000 \quad (23.2.1-7)$$

where $\lambda_C = \ln(2)/5,730$ years and $\lambda_R = \ln(2)/5,568$ years. t_C and t_R stand for the calibrated and the radiocarbon age, respectively. R_C and R_R are obtained from the measured $^{14}\text{C}/^{12}\text{C}$ ratio $R(t)$ which leads to

$$\Delta^{14}\text{C} = (e^{\lambda_C t_C - \lambda_R t_R} - 1) \cdot 1,000 \quad (23.2.1-8)$$

with the indexes c and r standing for calibrated ($\lambda = 5,730$ years) and radiocarbon ($\lambda = 5,568$ years), respectively. It is important to note that ages are reported as “calibrated years BP”, where “BP” stands for “before present” and “present” is by definition 1950.

The dating range of the radiocarbon method is basically determined by the ^{14}C half-life and the measurement errors. The limit is about 10 half-lives, i.e. about 60,000 years, which correspond to 1 permil of the present-day activity $(1/2)^{10}$. It should be noted that the uncertainties become considerably larger with increasing age, and there is no precise calibration curve yet beyond 12400 cal year BP.

There is also a limit on the younger side of the dating range. As a result of the combustion of fossil fuels (in which all the ^{14}C has decayed) the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio started to decrease at the beginning of the industrial age (see Fig. 23.2.1-1). This effect is called after its discoverer – the “Suess effect” (see Sect. 22.2). This decrease then stopped suddenly when the atomic bomb tests in the early 1950s injected large quantities of ^{14}C into the atmosphere. These rapid changes led to ambiguities which together with Maunder and Spoerer Minimum in solar activity make radiocarbon dating difficult for the first 400–500 cal year BP.

Radionuclides that potentially can fill partly or fully this gap of 400–500 years are the short-lived nuclides ^3H ($T_{1/2}$:12.4 years), ^{210}Pb ($T_{1/2}$:22.3 years), and ^{32}Si ($T_{1/2}$:145 years) (Fifield and Morgenstern 2009).

We now examine a direct consequence of the “Suess wiggles” in the calibration curve. Figure 23.2.1-2 shows two examples of the calibration process for steep and flat parts of the calibration curve, where the blue band represents the uncertainty. The first radiocarbon age of $2,650 \pm 40$ years (upper panel) leads to the

well-constrained calibrated age of $2,765 \pm 20$ years. If, however, we have a 200 year younger radiocarbon age with the same uncertainty of 40 years at 2,450 radiocarbon years, the resulting calibrated age is distributed over a large time window stretching from 2,700 to 2,350 years with a complicated non-Gaussian probability distribution. Improving the precision of the measurement to 10 years would not help in the slightest because there is a plateau in the calibration curve that introduces a great deal of ambiguity. Without additional information from samples with similar ages that lie off the plateau there is nothing one can do to improve the accuracy of the calibration (calendar) age. Such plateaus are formed by Grand Minima in which the production increases going back in time, thereby compensating for the decrease due to radioactive decay. So, while the Grand Minima excite a solar physicist, they are a pain for a radiocarbon person who would prefer the calibration curve to be a steadily decreasing straight line.

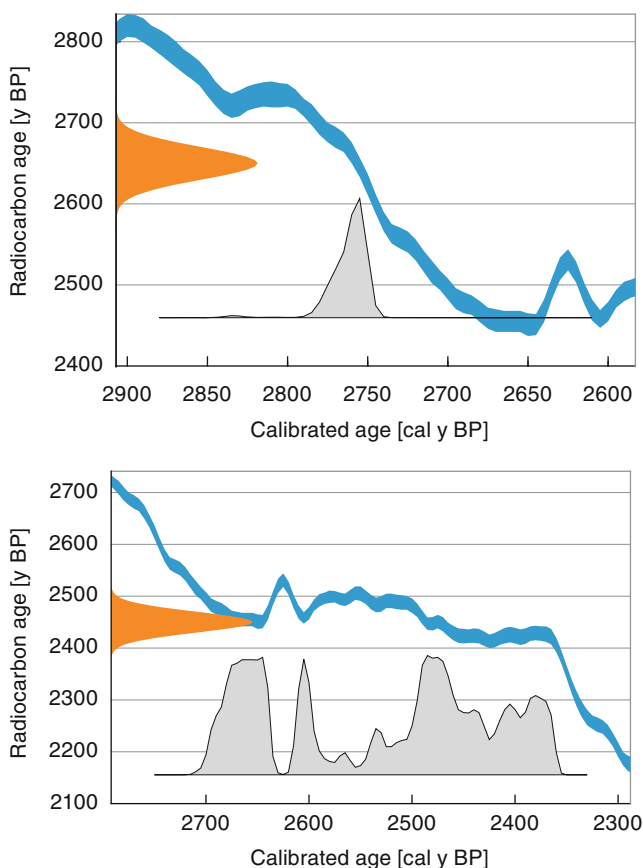


Fig. 23.2.1-2 Two examples of calibrating radiocarbon ages. The intersection of the Gaussian radiocarbon age distribution with the calibration curve (*blue band*) leads to very different probability distributions for the calibrated ages depending on the local form of the calibration curve

23.2.2 Exposure Dating

This is an example of “build-up dating” (Fig. 23.1-1) in which the production of radionuclides is switched on when a real system such as a rock is exposed to cosmic rays. The rock can be of lunar origin, a meteorite that is formed when a larger body in space breaks up into smaller bodies, or a rock on Earth which for some reason becomes exposed to cosmic rays at a certain time. A common example is that of terrestrial rocks that have been covered by a glacier for a long time, and which are then exposed to cosmic rays when the glacier retreats due to changes in the climate. In all cases it is important to know whether the rocks to be dated were free of the cosmogenic radionuclides of interest at the time the exposure began [i.e. that $C_0 = 0$ in (23.1-8)]. In cases where the system had already been exposed earlier to cosmic rays – for instance, if a glacier had already retreated earlier during a previous warm period – then things become more complicated and a combination of nuclides has to be applied.

The interaction of cosmic rays with the rock and the production of cosmogenic radionuclides in the rock are very similar to the interaction of cosmic rays with the atmosphere. The cosmic ray flux is attenuated with depth according to the mean free path Λ ($\sim 160 \text{ g cm}^{-2}$) and secondary particles are produced. This is particularly true for planetary rocks and meteorites. In the case of terrestrial rocks, the cosmic rays have already travelled through part of the atmosphere and the spectrum has been considerably modified. Depending on the altitude and the geomagnetic latitude, the production rates can be quite small. The two main differences compared to the interactions in the atmosphere are: (1) the elemental composition of rocks is different and (2) the cosmogenic radionuclides cannot move and mix as in the atmosphere. This property is very useful since it permits the study of the depth profiles of cosmogenic radionuclides and this provides the ability to test production models.

The concentration of the radionuclide at depth x in a rock is given by the equation (see Box 11.1.1):

$$N(x, t) = \frac{P_0 e^{-\rho x/\Lambda}}{\lambda} (1 - e^{-\lambda t}) \quad (23.2.2-1)$$

with

$N(x, t)$: the concentration (atoms g^{-1}) at depth x (cm) and time t

P_0 : the production rate (atoms $\text{g}^{-1} \text{ year}^{-1}$) at the top of the rock

ρ : the density in (g cm^{-2})

λ : the decay constant (y^{-1})

Λ : the mean free path (160 g cm^{-2})

Equation (23.2.2-1) describes an ideal case when the production rate P_0 is constant and the rock is a closed system (i.e., no radionuclides leave or enter the system). If P_0 is known and $N(x, t)$ is measured then the time t can be calculated (by taking the natural logarithm of (23.2.2-1):

$$t = -\frac{1}{\lambda} \ln \left(1 - \frac{\lambda N(x, t) e^{\rho x / \Lambda}}{P_0} \right) \quad (23.2.2-2)$$

In reality the production rate P_0 on Earth depends on altitude, latitude, and also time (as a consequence of solar and geomagnetic modulation of the cosmic radiation intensity: see Chap. 10). Furthermore, the axis of the geomagnetic dipole changes its position slowly relative to the rotation axis, leading to changes in the local cut-off rigidity and cosmic ray intensity (Sect. 5.8). However, since we are integrating the production over rather long periods of time, these short-term changes are not a serious problem as long as they do not affect the long-term average. A combination of different radionuclides with different half-lives can also be used to investigate the long-term constancy of the production rate $P_0(t)$. Even stable cosmogenic isotopes such as ^3He , ^{21}Ne , and ^{131}Xe are used for this purpose provided the rock is a closed system and no gases are lost.

Another complication is related to the fact that erosion of the exposed rock surface (inevitable on Earth) results in some of the produced cosmogenic radionuclides being lost. This can be expressed as (Box 11.1.1)

$$\frac{dN(x, t)}{dt} = P_0 e^{-\rho x / \Lambda} + \varepsilon \frac{dN(x, t)}{dx} - \lambda N(x, t) \quad (23.2.2-3)$$

with ε being the erosion rate (cm/y).

The differential equation (23.2.2-3) describes the time dependence of the concentration of the radionuclide $N(x, t)$ at depth x . The first term is the production rate $P = P_0 e^{-\rho x / \Lambda}$ which is attenuated with increasing depth. The second term is the loss caused by the erosion rate ε . This term is negative since the concentration N decreases with increasing depth. The last term is the radioactive decay.

If P and ε are constant the solution of the differential equation is:

$$N(x, t) = \frac{P_0 e^{-\rho x / \Lambda}}{\lambda + \varepsilon \rho / \Lambda} + \left(1 - e^{-(\lambda + \varepsilon \rho / \Lambda)t} \right) \quad (23.2.2-4)$$

This is the more general form of (23.2.2-1), in that it has the additional term $\varepsilon \rho / \Lambda$ describing the loss of atoms due to erosion, in addition to the loss due to radioactive decay. If $\varepsilon = 0$ Eq. (23.2.2-4) is identical to (23.2.2-1).

The fact that the erosion rate has the same effect as increasing the decay constant leads to lower equilibrium or saturation concentration at the rock surface ($x = 0$) for $t \rightarrow \infty$.

$$N_0 = \frac{P_0}{\lambda + \varepsilon \rho / \Lambda} \quad (23.2.2-5)$$

where N_0 is the saturation concentration at the rock surface.

Using $\mu = \rho/\Lambda$ (the mean free path expressed in centimetres) Eq. (23.2.2-5) becomes:

$$N_0 = \frac{P_0}{\lambda + \varepsilon\mu} \quad (23.2.2-6)$$

Taking natural logarithms of both sides of (23.2.2-4), the exposure time t is given by:

$$t = -\frac{1}{\lambda + \varepsilon\rho/\Lambda} \ln \left(1 - \frac{N(x,t)(\lambda + \varepsilon\rho/\Lambda)e^{\rho x/\Lambda}}{P_0} \right) \quad (23.2.2-7)$$

Most commonly, surface samples are used for exposure dating and (23.2.2-7) simplifies to

$$t = -\frac{1}{\lambda + \varepsilon\rho/\Lambda} \ln \left(1 - \frac{N_0(t)(\lambda + \varepsilon\rho/\Lambda)}{P_0} \right) \quad (23.2.2-8)$$

The main difficulty in exposure dating is not the depth dependence itself, but the necessity to know the erosion rate which is often not well known, and this means that we are left with one equation and two unknowns (age and erosion rate).

One solution to this problem is to find a rock with an exposure time $t \gg 1/(\lambda + \mu\varepsilon)$. In this case Eq. (23.2.2-6) applies and ε can be determined:

$$\varepsilon = (P_0/N_0 - \lambda)/\mu \quad (23.2.2-9)$$

Equation (23.2.2-6) shows that the steady state ratio N_0/P_0 is given by $1/(\lambda + \mu\varepsilon)$. This means that the precision of the determination of ε depends on the ratio $\lambda/\mu\varepsilon$. The larger λ is compared to $\mu\varepsilon$ the larger the uncertainty in the determination of ε .

The beauty of the cosmogenic radionuclide technique is that it may be possible to obtain additional solutions leading to increased accuracy through the use of two or more nuclides with different decay constants. A very common combination is ^{10}Be and ^{26}Al .

If the assumption of a constant erosion rate is correct, then $(P_0/N_0 - \lambda)$ must be invariant in a given rock for all nuclides [from (23.2.2-9)].

Figure 23.2.2-1 shows the calculated $^{26}\text{Al}/^{10}\text{Be}$ ratio in quartz. It is interesting to note that the ratio is bounded by the two limits of 6.1 and 3. The upper limit corresponds to the production ratio and prevails for ages younger than 10^4 years and high erosion rates. If the erosion rate is larger than 0.01 cm/y the measured ratio is always the production ratio independent of the age. In this case the erosion is so rapid that the produced atoms are removed before the faster decay of ^{26}Al ($T_{1/2}:0.73$ Myr) compared to ^{10}Be ($T_{1/2}:1.38$ Myr) can lower the ratio. The lower limit of 3 corresponds to the saturation ratio ($R_{\text{sat}} = R_{\text{prod}} \times \lambda_{\text{Be}}^{10}/\lambda_{\text{Al}}^{26}$) for a negligible erosion rate smaller than 10^{-6} cm year $^{-1}$. The curve of the ratio for $\varepsilon = 0$ sets a lower boundary for measurements of the $^{26}\text{Al}/^{10}\text{Be}$ ratio in quartz. Since the actual

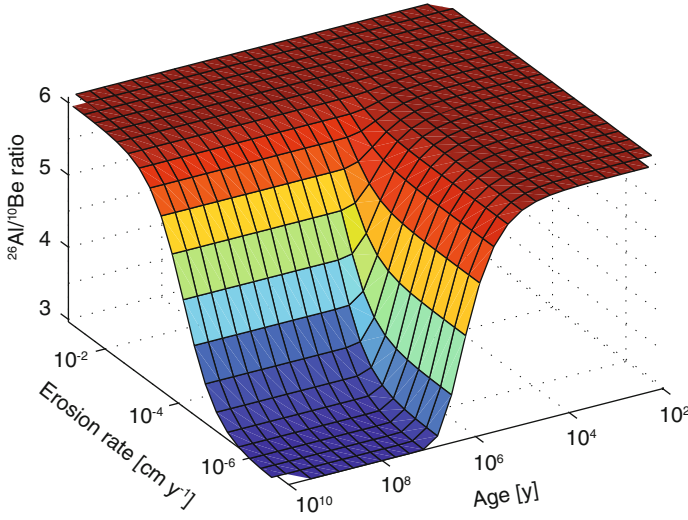


Fig. 23.2.2-1 $^{26}\text{Al}/^{10}\text{Be}$ ratio as a function of erosion rate and age in quartz. This ratio varies between two extreme values given by the production ratio of 6.1 and the saturation ratio of 3 for a negligible erosion rate. Ratios in between correspond to erosion rates in the range of 10^{-3} to 10^{-5} cm/y

measurements provide the ^{26}Al and ^{10}Be concentrations, an obvious way to present the data is to plot the $^{26}\text{Al}/^{10}\text{Be}$ ratio versus the ^{10}Be concentration. An upper boundary is obtained for $\varepsilon = 0$. Using (23.2.2-4) for $x = 0$, and $\varepsilon = 0$ leads to:

$$N(t) = (P_0/\lambda)(1 - e^{-\lambda t}) \quad (23.2.2-10)$$

$$R_{\text{Al/Be}} = \frac{N_{\text{Al}}}{N_{\text{Be}}} = \frac{P_{\text{Al}}\lambda_{\text{Be}}}{P_{\text{Be}}\lambda_{\text{Al}}} \frac{1 - e^{-\lambda_{\text{Al}}t}}{1 - e^{-\lambda_{\text{Be}}t}} \quad (23.2.2-11)$$

Replacing P_{Be} with:

$$P_{\text{Be}} = \lambda_{\text{Be}}N_{\text{Be}}/(1 - e^{-\lambda_{\text{Be}}t}) \quad (23.2.2-12)$$

leads finally to:

$$R_{\text{Al/Be}} = \frac{P_{\text{Al}}}{\lambda_{\text{Al}}N_{\text{Be}}(t)} \left[1 - \left(1 - \frac{N_{\text{Be}}\lambda_{\text{Be}}}{P_{\text{Be}}} \right)^{\lambda_{\text{Al}}/\lambda_{\text{Be}}} \right] \quad (23.2.2-13)$$

The combination of both boundaries is shown in Fig. 23.2.2-2.

All values for constant production and steady-state erosion rates must lie on this “island”. Each line in the island corresponds to a different erosion rate. As we know from Fig. 23.2.2-1 the lowest $^{26}\text{Al}/^{10}\text{Be}$ ratio of 3 corresponds to no erosion while

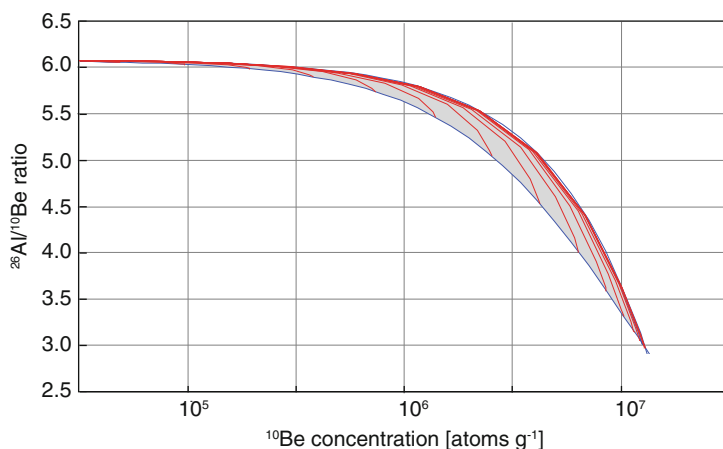


Fig. 23.2.2-2 “Island” of allowed $^{26}\text{Al}/^{10}\text{Be}$ ratios in quartz as a function of the ^{10}Be concentration in quartz

the ratio 6.1 means erosion rates greater than 0.01 cm/y. If experimental data do not lie on the “island” they are either wrong or the assumptions (constant production and erosion rates and simple single exposure history) are not fulfilled.

The production rate depends on the atmospheric depth (i.e. the altitude) of the exposed rock and the geomagnetic latitude, exactly as in the case of a neutron monitor. In fact, a rock with no erosion can be considered as a neutron monitor. The difference to the “atmospheric” neutron monitor is that the rock is integrating all the counts and does not therefore provide information about short-term temporal variations.

The local conditions may play a significant role in determining the production rate experienced by a rock surface. It may be surrounded by high mountains which shield it from part of the cosmic ray flux. Its surface may be covered with snow during the winter reducing the production rate even more. The production rate is not maximal if the surface is flat and horizontal. If the surface deviates from the horizontal, the strong dependence of cut-off rigidity (Fig. 5.8.2-3) on direction will have an important effect near the equator. Thus at the equator, the “East West” effect (Sect. 5.8) results in the cosmic ray flux from a 45° cone directed to the western horizon being 8.8 times greater than that from a similar cone towards the eastern horizon. Finally, it has to be kept in mind that the present-day conditions may not have pertained throughout the exposure time; for example, a rock may have rolled over at some time in the past. Careful and extensive sampling and the use of several cosmogenic radionuclides with different half-lives assist in avoiding most of these potential pitfalls. For practical determinations of exposure ages and erosion rates based on ^{10}Be and ^{26}Al we refer to Lal (1991) and Balco et al. (2008).

The first exposure ages were determined for meteorites. Exposed to the local interstellar spectrum without any attenuation due to the atmosphere and the geomagnetic field, the production rates and in particular the saturation concentrations of many cosmogenic radionuclides and stable noble gases were high enough to be

detected without accelerator mass spectrometry. On the other hand, corrections must be made for the ablation (i.e. erosion) of the outer surface during the meteorites' fiery entry into the atmosphere. These matters, and the use of the results from meteoritic and lunar material, are beyond the scope of this book. The interested reader is directed to specialist publications (Leya and Masarik 2009).

23.2.3 $^{10}\text{Be}/^{36}\text{Cl}$ - and $^7\text{Be}/^{10}\text{Be}$ -Dating

As noted previously, we need to know N_0 to use (23.1-4) to estimate the age of a sample. In theory, the use of the ratio of two cosmogenic radionuclides provides an elegant way to avoid this difficulty. The initial value of the ratio is then determined by nuclear physics (cross sections), and does not experience changes related to transport and deposition processes ("system effects") nor to production changes as long as the spectrum of the primary cosmic rays is constant. Production calculations show that even if the cosmic ray spectrum is modified by solar and geomagnetic modulation, the dependence on these changes is rather weak (see Sect. 10.3.3).

From (23.1-3) the general equation for the radioactive decay of a ratio is given by:

$$R_{a/b}(t) = \frac{N_a(t)}{N_b(t)} = \frac{N_a(0)}{N_b(0)} \frac{e^{-\lambda_a t}}{e^{-\lambda_b t}} = R_{a/b}(0) e^{-(\lambda_a - \lambda_b)t} = R_{a/b}(0) e^{-\lambda_{a/b} t}$$

This shows that the ratio changes with a new decay constant $\lambda_{a/b} = \lambda_a - \lambda_b$ and a new half-life

$$T_{a/b} = \frac{T_a T_b}{T_a + T_b}. \quad (23.2.3-1)$$

Inserting $R_{a/b}(t)$ and $\lambda_{a/b}$ into (23.1-4) gives the date estimate as

$$\tau = \frac{1}{\lambda_{a/b}} \ln \left(\frac{R_{a/b}(0)}{R_{a/b}(t)} \right) \quad (23.2.3-2)$$

And in a manner analogous to the derivation of (23.2.1-6), the dating uncertainty $d\tau$ depends on the relative uncertainty of R and the half-life:

$$d\tau = \frac{1}{\lambda_{a/b}} \frac{\Delta(R_{a/b}(0))}{R_{a/b}(0)} = \frac{\Delta(R_{a/b}(0))}{R_{a/b}(0)} \frac{T_{a/b}}{\ln(2)} \quad (23.2.3-3)$$

In the following, we consider dating using the two ratios, $^{10}\text{Be}/^{36}\text{Cl}$ and $^{10}\text{Be}/^7\text{Be}$. From (23.2.3-1) the half-lives of these ratios are 386 kyr and 53.2 days, respectively. The latter is unchanged from the half-life of ^7Be (53.2 days) because the half-life of ^{10}Be ($T_{1/2} = 1.38 \text{ My}$) is so much greater and the two terms

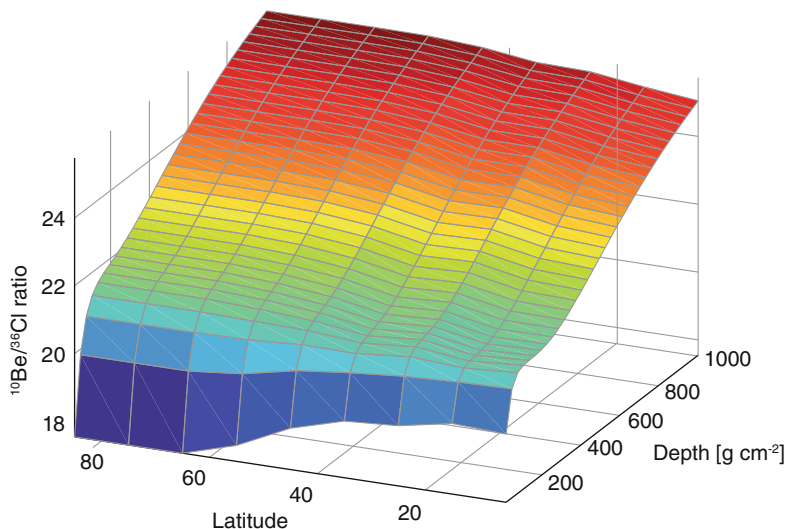


Fig. 23.2.3-1 Dependence of the $^{10}\text{Be}/^{36}\text{Cl}$ production ratio for the present geomagnetic field intensity and a solar modulation function $\Phi = 550$ MeV according to Masarik and Beer (1999). The altitude dependence is mainly due to the contribution from the reaction $^{36}\text{Ar}(n,p)^{36}\text{Cl}$

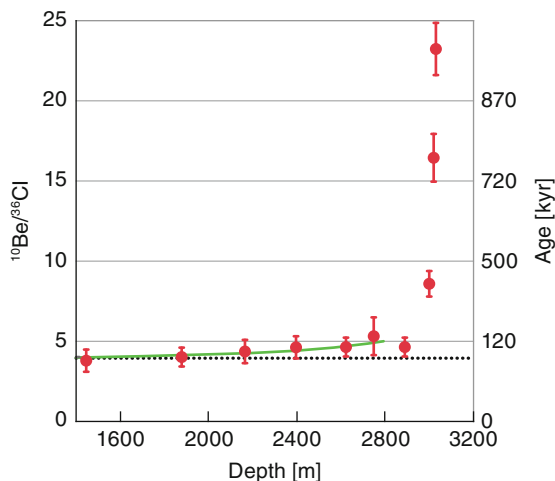
involving it cancel out in (23.2.3-1). Note that in both cases the half-life of ^{10}Be is greater than that of the radionuclide in the denominator, and as a consequence, the decay constant of the ratio, $\lambda_{a/b} = \lambda_a - \lambda_b$, is negative. Therefore both ratios increase with time. This is obvious from first principles, since the term in the denominator decays faster than that in the numerator.

An important pre-condition for the use of these two ratios is that the transport and scavenging processes (from the point of production in the atmosphere) as well as any process in the ice should be equal for both nuclides. While this condition is fulfilled in the case of $^{10}\text{Be}/^7\text{Be}$, this is not always the case for $^{10}\text{Be}/^{36}\text{Cl}$.

The $^{10}\text{Be}/^{36}\text{Cl}$ production ratio depends considerably on altitude (Fig. 23.2.3-1) due to the relatively large cross section of the reaction $^{36}\text{Ar}(n,p)^{36}\text{Cl}$ for neutron energies around 10 MeV. The dependence on latitude, geomagnetic field intensity, and solar modulation function (Φ) are comparatively small ($<10\%$).

The fact that the $^{10}\text{Be}/^{36}\text{Cl}$ ratio grows with a half-life of 386 kyr seems ideally suited for dating polar ice cores, which can reach ages of about one million years. However, the first measurements using polar ice showed surprising results. The ratios for relatively young samples were in the range 5–10 which was considerably smaller than that expected (~ 20 according to Masarik and Beer 1999) and even worse, they were far from constant. The discrepancy between the measured and the calculated ratio may be related to uncertainties in the cross sections and is not a serious problem for dating. However, the fluctuation of the ratios for samples with similar ages poses a big problem. An error of 10% in the ratio corresponds to an error of 54 ky in age. In most cases, errors larger than 20% are too large to provide any useful ages.

Fig. 23.2.3-2 $^{10}\text{Be}/^{36}\text{Cl}$ dating of the lowermost 600 m of the GRIP ice core from Greenland (Willerslev et al. 2007). The *green line* represents the known depth-age model



The explanation for these fluctuations is related to the different geochemistry of Be and Cl. Part of ^{36}Cl is deposited as H^{36}Cl . During the recrystallization processes in the firn (the first ~ 100 years after deposition as snow) HCl can be released and it then diffuses towards the atmosphere. Some H^{36}Cl is adsorbed again before reaching the surface, the rest is lost. This process leads to a larger $^{10}\text{Be}/^{36}\text{Cl}$ ratio and therefore to an overestimation of the ages. The errors are small or even negligible if the snow accumulation rate is large and if the dust content in the snow is relatively high. In the latter case, HCl reacts with the dust forming salts. More details about the loss of ^{36}Cl were given in Sect. 21.3.2.

Based on this understanding of the ^{36}Cl loss, careful selection of samples with high dust content enables the application of the $^{10}\text{Be}/^{36}\text{Cl}$ dating method, although with lower accuracy than was originally hoped. Figure 23.2.3-2 shows the result of the analysis of the lowermost 600 m in the GRIP ice core from Greenland. The green line shows the known depth-age model. The dotted line represents the mean $^{10}\text{Be}/^{36}\text{Cl}$ ratio in the top part of the ice core. Note the rapid age increase towards the bedrock at 3,028 m depth as a consequence of ice flow thinning the annual layers.

The advantage of the $^{10}\text{Be}/^7\text{Be}$ ratio is that both nuclides are chemically identical and no problems such as in the case of $^{10}\text{Be}/^{36}\text{Cl}$ are to be expected. The short half-life of 53.2 days makes this ratio an ideal tracer to study atmospheric transport processes.

The $^{10}\text{Be}/^7\text{Be}$ ratio as a function of altitude and latitude is shown in Fig. 23.2.3-3. Again the ratio is strongly altitude dependent on the lowest ratios occurring at the top of the atmosphere. However, this time the reason is that the $^{14}\text{N}(p,nx)^7\text{Be}$ cross section has a maximum at low energies (see Fig. 10.3.2-1). As a consequence the $^{10}\text{Be}/^7\text{Be}$ ratio varies by a factor of two with a mean value of 0.7. The dependence on the latitude and the geomagnetic and solar modulation is comparatively small.

The application of the $^{10}\text{Be}/^7\text{Be}$ ratio for studies of atmospheric transport was discussed in Sect. 19.2.

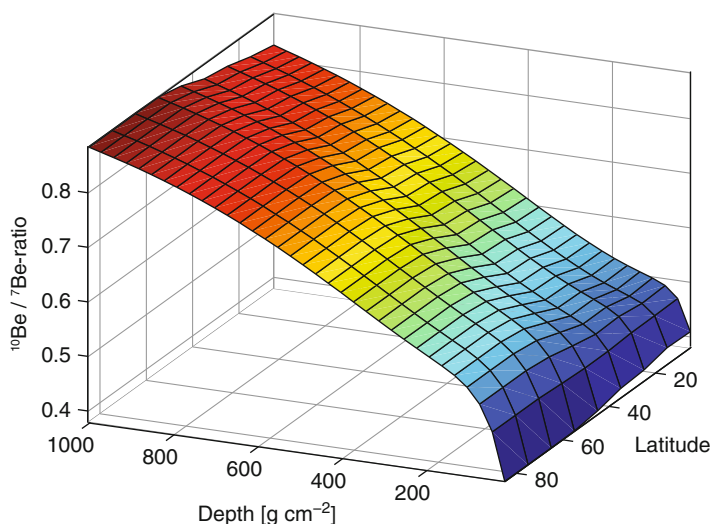


Fig. 23.2.3-3 Dependence of the $^{10}\text{Be}/^{7}\text{Be}$ production ratio for the present geomagnetic field intensity and a solar modulation function $\Phi = 550$ MeV according to Masarik and Beer (1999, 2009). The strong altitude dependence is due to the large cross section of the reaction $^{14}\text{N}(p,nx)^7\text{Be}$ at low energies

23.3 Synchronization of Records

In the previous sections, we made use of radioactive decay to provide absolute dates for samples or records. In this section, we describe methods used to transfer a known time scale from record A to record B which does not have a precise time scale. This procedure is called synchronization or relative dating and is discussed, in part, in Chap. 14. As we will see, the records to be synchronized can be either continuous or discrete signals. It is very important to make sure, however, that the signals used for synchronization are really simultaneous. To illustrate this we use the example of a $\delta^{18}\text{O}$ signal continuously measured in an ice core from Dome C in Antarctica (Fig. 14.4-1). Since $\delta^{18}\text{O}$ is a proxy for temperature it reflects climate change, and we can reasonably assume that similar climate changes have occurred in all Antarctic ice cores at about the same time. Even if we do not know the absolute time of the climate changes the records can be synchronized on a relative time scale. However, it may be dangerous to synchronize Antarctic $\delta^{18}\text{O}$ records with $\delta^{18}\text{O}$ records from Greenland, say. For example, a reduction of the thermohaline circulation may reduce the heat transport from the south to the north and a cooling episode in the north could then be coincident with a warming episode in the south. In this case synchronizing $\delta^{18}\text{O}$ from the two hemispheres would be clearly wrong.

A continuous signal such as $\delta^{18}\text{O}$ is ideal for synchronization. However, discrete signals, often called time markers, are useful as well. Volcanic eruptions are a good example – they inject huge amounts of sulphates and other chemical constituents into the stratosphere. There they are mixed globally [in a manner similar to the

cosmogenic radionuclides (see Chapter 13)] and then precipitated to Earth after a residence time of 1–2 years, resulting in sharp simultaneous peaks in ice cores. Volcanic eruptions occur all the time (see Fig. 14.3-3) and many have occurred during the past millennium. Many have been well documented at the time of occurrence, and therefore we know their absolute dates of occurrence. Earlier volcanic events can be used for relative dating. The debris from large eruptions in the equatorial zone can be found in ice cores in both hemispheres. It should be noted that volcanic events occur relatively frequently, leading to the possibility of misinterpretation and mixing up eruptions which occurred at slightly different times.

In the following, we provide an example where worldwide changes in the production of the cosmogenic radionuclides are used for synchronization.

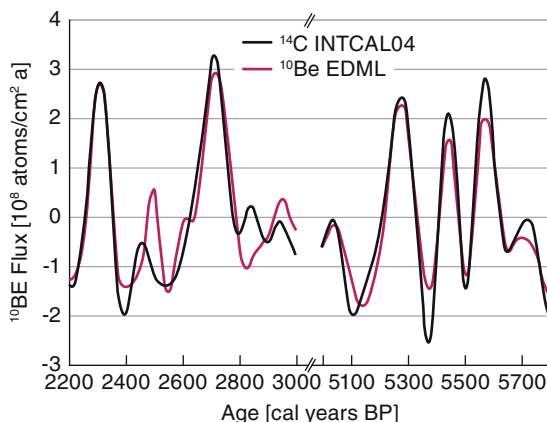
23.3.1 ^{10}Be or ^{36}Cl with ^{14}C During the Holocene

A good example of synchronization is provided by the transfer of the well-known time scale of the ^{14}C record obtained from well-dated tree rings, to the ^{10}Be record in ice cores. As discussed in Chap. 10 ^{10}Be and ^{14}C are produced simultaneously in the atmosphere by cosmic rays. The production rate is modulated by solar activity, and this modulation signal can be used for synchronization. The absolute time scale is then provided by the ^{14}C record from tree rings. By dating many trees that overlap in time, a dendrochronology can be established which is not limited by the age of the trees but rather by the availability of fossil trees. Today the European dendrochronology (mainly based on oak and fir trees) goes all the way back to about 12000 BP, when the first trees appeared after the last glacial epoch. An extension of the record back to about 40000 years BP (the limit of the radiocarbon method) must be based on a new dendrochronology. Work is in progress to provide this using Kauri trees in New Zealand.

Synchronization of ^{10}Be with ^{14}C is not straightforward. The problem is that the measured signals do not directly reflect the production signal, as a consequence of the transport and exchange processes within the carbon system (see Sect. 13.5.3). While the ^{10}Be signal is delayed by only about 1 year the ^{14}C lags several years and its amplitude is strongly attenuated, depending on how quickly the production changes (see Fig. 13.5.3.2-2). Therefore, proper synchronization requires that both of the measured signals are converted into their respective production signals. To this end, ^{10}Be can be simply shifted backwards by 1 year, while in the case of ^{14}C a carbon cycle model is needed to perform this task (see Sect. 13.5.3).

As an example, Fig. 23.3.1-1 shows a comparison of ^{10}Be (red line) from Dronning Maud Land in Antarctica with the ^{14}C production (black line) rate derived from INTCAL04 for two time intervals of 800 years each (Ruth et al. 2007). Both records have been band filtered (100–1,000 years). The figure illustrates that large peaks caused by grand solar minima can be synchronized unambiguously. Small peaks may be related to system effects and have to be treated with caution.

Fig. 23.3.1-1 Two time intervals for which the ^{10}Be records from Dronning Maud Land in Antarctica were synchronized with the ^{14}C production derived from the INTCAL04 calibration curve. Note break in time scale. After Ruth et al. (2007)



This raises the question of the accuracy of this method. The Sun has modulated cosmogenic radionuclide production on time scales ranging from 11 to 2,300 years. Theoretically, it would be possible, although very difficult and extremely expensive, to correlate the two records within a few years. In practice, an accuracy of 10–20 years is more realistic given the fact that the majority of the ^{14}C measurements have been performed on 5–10 consecutive combined tree rings. However, with the development of the AMS technique the number of annual ^{14}C records is steadily growing, providing the possibility to date particularly interesting periods in the far past with annual accuracy.

Synchronization between ^{10}Be and ^{14}C is probably the most accurate dating technique for ice cores over the past 14,000 years. However, it requires continuous high-resolution records of both nuclides. The ^{14}C record is the basis for the radiocarbon dating method (see Sect. 23.2.1 and Fig. 23.2.1-1) and therefore is well known; however, the establishment of a detailed ^{10}Be record for each ice core requires great effort and it is not feasible for dating purposes alone.

23.3.2 The Use of Time Markers

While relative synchronization of continuous records is independent of any preliminary time scale, time markers can only be used to improve a provisional time sale. This is especially true when the number of time markers is large and it is not a priori clear which marker in one record belongs to several markers in the other record.

Although not as distinct as volcanic eruptions, cosmogenic radionuclides can also be used as time markers. In principle, every grand solar minimum can be considered as a time marker as can be seen in Figs. 23.2.1-1 and 23.3.1-1. Looking to the future, the grand solar minima may be particularly useful in dating the ^{10}Be record in the glacial epoch that extends far into the past beyond the existing ^{14}C record, and particularly for times beyond the potential limit of the ^{14}C record.

Finally, we mention briefly two other types of radionuclide time marker which are of importance.

Bomb Pulse. As discussed (Sect. 19.3), the nuclear bomb tests that occurred between 1950 and 1963 introduced great quantities of ^3H , ^{14}C , and ^{36}Cl , and also ^{137}Cs and ^{90}Sr into the atmosphere. These bomb peaks are excellent modern markers which are widely used in dating sediments and ice cores. Combined with ^{137}Cs released during the Chernobyl accident in 1986, they provide the means to measure the sedimentation rates, throughout the world, between 1963 and 1986.

Geomagnetic Events. As with the grand solar minima, large geomagnetic events increase the mean global cosmogenic radionuclide production rate by about a factor of 2 (Fig. 10.3.3-3). However, they are very rare and less distinct because they occur on millennial time scales (see Sect. 21.2). In fact, only the Laschamp excursion has been used successfully to synchronize Antarctic and Greenlandic ice cores (Beer et al. 1992; Raisbeck et al. 2007). It has been dated by $^{40}\text{Ar}/^{39}\text{Ar}$, K-Ar, and $^{238}\text{U}/^{230}\text{Th}$ to $40,650 \pm 950$ years (Singer et al. 2009) (see Sect. 21.2).

References

- Balco G, Stone JO, Lifton NA, Dunai TJ (2008) A complete and easily accessible means of calculating surface exposure ages or erosion rates from Be-10 and Al-26 measurements. *Quat Geochronol* 3(3):174–195
- Beer J, Johnsen SJ, Bonani G, Finkel RC, Langway CC, Oeschger H, Stauffer B, Suter M, Wölfli W (1992) ^{10}Be peaks as time markers in polar ice cores. Paper presented at Proceedings of the Erice Nato Workshop. The last delegation: absolute and radiocarbon chronologies. Springer, Berlin
- Fifield LK, Morgenstern U (2009) Silicon-32 as a tool for dating the recent past. *Quat Geochronol* 4(5):400–405
- Lal D (1991) Cosmic-ray labeling of erosion surfaces – insitu nuclide production-rates and erosion models. *Earth Planet Sci Lett* 104(2–4):424–439
- Leya I, Masarik J (2009) Cosmogenic nuclides in stony meteorites revisited. *Meteorit Planet Sci* 44(7):1061–1086
- Masarik J, Beer J (1999) Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res* 104(D10):12,099–12,111
- Masarik J, Beer J (2009) An updated simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J Geophys Res-Atmos* 114:D11103
- Raisbeck GM, Yiou F, Jouzel J, Stocker TF (2007) Direct north-south synchronization of abrupt climate change record in ice cores using Beryllium 10. *Clim Past* 3:541–547
- Reimer PJ et al (2004) IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–1058
- Ruth U et al (2007) “EDML1”: a chronology for the EPICA deep ice core from Dronning Maud Land, Antarctica, over the last 150 000 years. *Clim Past* 3(3):475–484
- Singer BS, Guillou H, Jicha BR, Laj C, Kissel C, Beard BL, Johnson CM (2009) Ar-40/Ar-39, K-Ar and Th-230-U-238 dating of the Laschamp excursion: a radioisotopic tie-point for ice core and climate chronologies. *Earth Planet Sci Lett* 286(1–2):80–88
- Stuiver M, Polach HA (1977) Discussion reporting of ^{14}C data. *Radiocarbon* 19(3):355–363
- Willerslev E et al (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* 317:111–114

Glossary

This glossary is restricted to selected words and phrases that arise in a number of different chapters in the book, or are deemed to be sufficiently unusual and important to merit definition.

Accelerator Mass Spectrometry (AMS) An extremely sensitive technique that determines the mass number of individual atoms, allowing isotopic ratios of a sample to be determined.

Anthropogenic Manmade contribution to the environment.

Atmospheric depth A convention where position in the atmosphere is measured from the top in units of g cm^{-2} .

Barn A unit of area (10^{-24} cm^2) used to quantify the probability of a nuclear reaction.

Before Present (BP) A convention where time is stated relative to the year 1950 AD.

Binding energy The energy required to remove a neutron or proton from an atomic nucleus.

Box model A simple mathematical model that can be used to compute the temporal and spatial behavior of a total system (e.g., the carbon cycle), and allows perturbations to be evaluated.

Calibrated age The “true” age estimate derived from the radiocarbon age (See below) and a contemporary calibration curve.

Carbon cycle The complex exchange of CO_2 between atmosphere, biosphere, and the oceans.

Cascades, nucleonic, proton, or neutron A “chain reaction” wherein a single cosmic ray initiates a “shower” of particles throughout the atmosphere.

Cosmic ray anomalous Low energy cosmic rays produced by photo-ionization in the heliosphere.

Cosmic ray differential energy spectrum (or intensity) The manner in which the intensity of the cosmic radiation depends upon particle energy.

Cosmic ray galactic and solar Galactic and solar being produced in the galaxy, and sun, respectively.

Cosmic ray propagation equation Specifying the effects of the solar wind and heliospheric magnetic fields upon the intensity of the cosmic radiation.

Cross section (excitation function) A parameter that specifies the probability that a nuclear reaction will occur.

Current sheet Surface where the solar magnetic field changes its polarity.

Cyclotron Particle accelerator also used for AMS of the noble gasses Ar and Kr.

Decay constant A parameter characterizing the rate of decay of a radioactive nuclide. Inversely related to the half-life.

Dendrochronology Dating based upon counting back the annual growth rings in trees.

Deposition, wet, dry, and gravitational The several processes that transfer the cosmogenic radionuclides from the atmosphere to the surface of earth.

Excitation function The dependence of a nuclear cross-section on energy.

Firn Partially consolidated snow before it converts to solid ice.

Fluence A measurement of the radiation received during a given period of time.

General Circulation Models (GCM) Numerically intensive computer models that simulate the three-dimensional circulation and interchange effects in the atmosphere.

Geomagnetic cut-off The lower limit to the cosmic ray energy (or rigidity) that can reach a specified point on earth.

Geomagnetic field The magnetic field of the earth.

Geomagnetic modulation Change in the production rate of the cosmogenic radionuclides induced by the geomagnetic field.

Gleissberg cycle The ~87-year cycle in solar magnetic activity.

Grand minimum A period from 50 to 100 years in duration when the Sun is relatively inactive (few sunspots).

Hale cycle The 22-year cycle in solar magnetic fields.

Half-life The period of time in which 50% of a radioactive nuclide will decay. Inversely related to the decay constant.

Hallstatt cycle The ~2300-year periodicity in solar activity.

Heliosheath The region of the heliosphere beyond the termination shock (See Termination shock).

Heliosphere The region of space under the control of the Sun.

Holocene The 11,600-year period from the end of the last glacial epoch to the present.

Ionization The production of electrons and positive ions by a charged particle.

Ionization chamber Early form of recording cosmic ray detector, 1933–1979.

Laschamp event A major decrease (excursion) in the strength of the geomagnetic field ~41000 BP.

Local Interstellar Spectrum (LIS) The cosmic ray spectrum in our part of the galaxy- that is, outside the heliosphere.

Loess Sedimentary layers formed from wind-blown dust.

Magnetohydrodynamics Study of the properties of magnetized plasma.

Mean life time The inverse of the Decay constant.

Modulation function (potential) Parameter used to quantify the degree of modulation of the cosmic radiation intensity by solar activity.

Monte Carlo techniques Mathematical techniques that use probabilistic concepts to model physical processes such as the effects of charged particles as they pass through matter.

Neutron monitor The most important form of ground level cosmic ray recording instrument.

Parker spiral field The interplanetary magnetic field convected outwards by the solar wind.

Pitch angle The angle between the velocity vector of a cosmic ray and the magnetic vector.

Principle components analysis A mathematical technique used to extract a common signal that is present in a number of different data records.

Radiocarbon age A first approximation to the age of a sample based on the decay equation (See also Calibrated Age).

Relativistic Relating to particles moving at velocities close to that of light.

Response function A mathematical representation of the relative contributions to the counting rate of a cosmic ray instrument, or to a cosmogenic measurement.

Rosetta stone As used here, a combination of a number of scientific techniques that allow us to interpret the cosmogenic radionuclides in terms of the cosmic radiation intensity in the vicinity of Earth.

Schwabe cycle The 11-year solar cycle (in particular, the 11-year sunspot cycle).

Solar modulation Variations of the cosmic ray intensity, and of the production of the cosmogenic radionuclides, as a consequence of changes in solar activity.

Spaceship Earth A network of super neutron monitors located such that they provide directional viewing over the whole celestial sphere.

Spallation A nuclear process that disrupts the atomic nucleus generating nuclei of lower atomic number.

Specific yield function A mathematical representation of the energy sensitivity of various cosmic ray instruments and cosmogenic measurements.

Stoermer cut-off The minimum energy (rigidity) cosmic ray that can reach the top of the atmosphere.

Stratosphere The thermally stratified atmospheric layer above the troposphere between altitudes of ~10 and 50 km.

Suess effect The steady reduction in the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio after the mid nineteenth century due to the increasing combustion of fossil fuel.

Super neutron monitor Widely used cosmic ray detector, 1965-onwards.

Supernova The violent explosion of a star (quite rare).

System effects Modifications to the cosmogenic fluxes due to processes in atmosphere, ocean, etc.

Tandem accelerator A particle accelerator commonly used in accelerator mass spectrometry.

Termination Shock The abrupt discontinuity between the supersonic solar wind in the inner heliosphere and the subsonic material in the Heliosheath.

Thermohaline circulation Large scale ocean circulation driven by solar energy.

Time Before Present (BP) A convention where time is measured relative to the year 1950 (defined as present).

Tropopause The boundary between the stratosphere and the troposphere (See Troposphere and Stratosphere).

Troposphere The lowest convective portion of the atmosphere, to heights of 7–19 km (See Stratosphere).

Troposphere exchange The complex process whereby stratospheric air gains access to the troposphere, primarily at mid-latitudes.

Index

A

Accelerator, 283
Accelerator mass spectrometry (AMS), 7, 17–18
Accumulation, 182
ACR. *See* Anomalous cosmic radiation (ACR)
Acrotelm, 258
Aeolian, 370
Air shower array, 30
Albedo, 3, 206
Alchemists, 136
Aletsch Glacier, 322
Allochthonous material, 252
Alpha, 142
AMS. *See* Accelerator mass spectrometry (AMS)
Anomalous cosmic radiation (ACR), 334
Aphelion, 187
Archaeology, 8
Archeomagnetism, 66
Archimedes spiral, 41
Astronomical unit (AU), 26, 187
Astrophysics, 8
Atmosphere, 8
AU. *See* Astronomical unit (AU)
Aurora australis, 75
Aurora borealis, 75
Auroral ovals, 76
Avogadro number, 144

B

Barn, 144
Berkner island, 348
Bethe-Bloch formula, 148
Bioavailability, 393
Biokinetics, 394

Biophysics, 8
Biosphere, 8
Bioturbation, 250
B/M reversal, 376
Bow shock, 42
Bremsstrahlung, 193
Brewer–Dobson circulation, 209

C

Carrier, 286
Catotelm, 258
Cheltenham, 96
Chernobyl, 248, 341
Christchurch, 96
Chromatographic column, 381
Climax, 95
CMEs. *See* Coronal mass ejections (CMEs)
Colatitude, 65
Coriolis force, 206
Coronal mass ejections (CMEs), 38, 39
Coulomb barrier, 149
Coulomb-interactions, 148
Cross-sections, 143
Current sheet, 41–43
Cut-off rigidity, 152

D

Dalton, 95, 300
Dansgaard–Oeschger event, 352
Decay chain, 196
Dendro-age ratio, 404
De Vries cycle, 298
Discrete Fourier transform (DFT), 299
Distribution coefficient (K_d), 381
Diurnal variation, 109

$\delta^{18}\text{O}$, 246

Dome C, 249

Dronning Maud Land, 416

E

East–West effect, 71–72

ECHAM5-HAM, 349

Eddy diffusivity, 228

Ejecta, 63

Eluted downward, 381

Environmental system, 14

EPICA. *See* European Project for Ice Coring in Antarctica (EPICA)

Epithermal neutron flux, 184

Erosion, 182

European Project for Ice Coring in Antarctica (EPICA), 377

Evapotranspiration, 355

Exosphere, 205

Exposure age, 181–182

Extratropical pump, 209

F

Faraday cup, 284

Ferrel cell, 212

Fiescherhorn, 348

Fluence, 325

FLUKA, 158

Fluvial, 370

Forbush decrease, 62

Fourier spectrum, 303

Fourier transform, 299

Frequency domain, 302

Fukushima, 341

Fulvic substances, 381

G

Gamma ray burst, 337

Gauss, 36

GCM. *See* General circulation model (GCM)

GEANT, 94, 150

Geiger counter, 111

General circulation model (GCM), 216

Geochemical cycles, 14

Geomagnetic field, 90

Geomorphological process, 370

Geophysics, 8

Geosphere, 8, 369

GLE. *See* Ground level events (GLE)

Glacisberg, 298

Global inventory, 171

Globally merged interaction regions (GMIR), 106–109

Godhavn, 96

GOES, 126

Grand Minimum, 151

Grenzgletscher, 348

Ground level events (GLE), 112

Ground water, 359

Gulyia, 348

Gyroradius, 23, 25–26

H

Hadronic interactions, 150

Hale cycle, 47, 312

Half-life, 7, 171

Hallstatt cycle, 299

Hallstatt period, 229

Hectopascal, 141

Heliographic latitudes, 36

Heliomagnetic cycle, 101

Heliomagnetic magnetic field (HMF), 57, 317, 324

Heliomagnetograph, 317

Heliopause, 42

Helioseismology, 3

Heliosheath, 42

Heliosphere, 11, 41–43, 102

Heliospheric phenomena, 36

Hermanus, 95

Heterodyne frequency, 304

HMF. *See* Heliomagnetic magnetic field (HMF)

Holocene, 311

Horse latitudes, 207

Huancayo, 96

Huascarán, 348

Humic substances, 381

Humification, 258

Hydrology, 8

Hydrosphere, 8

I

IMP. *See* Interplanetary monitoring platform (IMP)

In-situ production, 180

Insolation, 207

Interplanetary monitoring platform (IMP), 108

Intertropical convergence zone (ITCZ), 207

Ionization chamber, 80–83

Ionosonde, 111

ITCZ. *See* Intertropical convergence zone (ITCZ)

K

Kaons, 160
 K_d . *See* Distribution coefficient (K_d)
 Kepler, 339
 Kiloparsec, 332

L

La Hague, 341, 394
 Large Hadron Collider (LHC), 200
 Laschamp event, 232
 Laschamp geomagnetic excursion, 248
 Lepton, 192
 LHC. *See* Large Hadron Collider (LHC)
 Libby, Williard, 401
 Light year, 26
 Liquid scintillation counters (LSC), 279
 LIS. *See* Local interstellar spectrum (LIS)
 Lithosphere, 179
 Local interstellar spectrum (LIS), 29, 44, 334
 Loschmidt number, 144
 LSC. *See* Liquid scintillation counters (LSC)

M

Magnetohydrodynamic waves, 41
 Magnetosphere, 73–76
 Maunder, 95
 Maunder Minimum, 205
 Mean free path, 144
 Meridional transport, 208
 Meson(ic) component, 142, 143
 Mesosphere, 205
 Methanogenic bacteria, 258
 Methylchloride, 236
 Milankovitch cycles, 312
 Millibar, 141
 Modulation function, 52
 Modulation potential, 52
 Monte Carlo code, 150
 Monte Carlo techniques, 13
 Mount Palomar, 8
 Mount Tambora, 247
 μ -meson, 143
 Muon, 100
 Muon telescope, 80–83

N

Nanotesla, 28
 Neutral sheet, 43
 Neutrino, 192
 Neutron monitor, 83–88

Nevada, 348
 Novaja Semlya, 348
 Novi Sibirsk, 348
 Nucleonic component, 142
 Nucleosynthesis, 195
 Nutation, 312

O

Octupole, 66
 Oort, 310

P

Paleomagnetic method, 67
 Paleontology, 325
 Paleothermometer, 390
 Parker spiral field, 37
 Parsec, 26
 Pedosphere, 214
 Photospallation, 337
 Photosphere, 41
 Pions, 143
 Pitch angle, 127
 Plastic scintillator, 281
 Plumes, 36
 Poisson distribution, 287
 Polar aurora, 73–76
 Power spectra, 302
 Primordial radionuclides, 195

Q

Quadrupole, 66

R

Radioactive decay, 182
 Remanent magnetization, 375
 Residence time, 273
 Response function, 91
 Resuspension, 250
 Rigidity, 23, 25–26
 Riometer, 111

S

Saturation concentration, 182
 Schnidejoch, 322
 Schwabe, 312
 Scintillators, 88
 SCR. *See* Solar cosmic rays (SCR)
 Sellafield, 341, 394

SEP. *See* Solar energetic particle (SEP)

Sinusoidal, 345

Solar cosmic rays (SCR), 187

Solar dynamo, 310

Solar energetic particle (SEP), 112

Solar flares, 39

Solar wind, 37

Space science, 8

Spallation, 332

Speciation, 381

Specific yield, 172

Speleothems, 253–254

Spike, 286

Spoerer, 95, 300

Spoerer Minimum, 76

STE. *See* Stratosphere-troposphere exchange (STE)

Stoermer cut-off, 70, 90

Stoke's law, 214

Stopping power, 149

Stratosphere, 141

Stratosphere-troposphere exchange (STE), 345

Streamers, 36

Subduction, 225, 370

Suess cycle, 298

Suess effect, 231

Sunspot number, 34–40

Sunspots, 3

Supernova(e), 17, 27

T

Tandem accelerator, 284

Tauon, 192

Tephra, 247

Termination shock, 41–43

Tesla, 36

THC. *See* Thermohaline circulation (THC)

Thermal neutron flux, 184

Thermohaline circulation (THC), 227, 352, 390

Thermosphere, 205

Thyrototoxicosis, 394

Tien Shan, 348

Time domain, 303

Time of flight, 284

Transdisciplinary, 5

Transfer function, 267

Tritium, 359

Tropopause, 141

Troposphere, 141

Tycho, 339

U

Ultraviolet (UV), 3

Utah, 348

UV. *See* Ultraviolet (UV)

V

Varves, 246

Vela Junior, 338

Voyager, 41

W

Wavelet analysis, 302

Wolf, 310

X

X-rays, 3, 39

Y

Younger Dryas (YD), 231, 391